

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2002, 20:40:02 ; Search time 51 Seconds  
(without alignments)  
772.845 Million cell updates/sec

Title: US-09-699-652A-14  
Perfect score: 2194  
Sequence: 1 MAMAGHAPGGALPLILLVWS.....DDVYVDLIRFLRENGWHNSY 410

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1055	48.1	344	2 E84526	probable lysosomal
2	557.5	25.4	403	2 T33198	hypothetical prote
3	553	25.2	397	1 JC4017	triacylglycerol li
4	549	25.0	405	2 T22675	hypothetical prote
5	546	24.9	398	2 S07145	triacylglycerol li
6	539.5	24.6	395	1 LIRTT	triacylglycerol li
7	523	23.8	405	2 H88930	protein R1G11.14
8	515.5	23.5	399	2 S41408	lysosomal acid lip
9	513	23.4	399	2 G01416	lysosomal acid lip
10	500.5	22.8	426	2 T20480	hypothetical prote
11	474	21.6	411	2 T22290	hypothetical prote
12	466.5	21.3	411	2 G89074	protein K04A6.5 [i
13	416.5	19.0	559	2 J70949	egg-specific prote
14	392.5	17.9	1585	2 T31611	hypothetical prote
15	356	16.2	443	2 T39540	triglyceride lipas
16	345.5	15.7	467	2 T41053	triglyceride lipas
17	309	14.1	460	2 T39443	probable triglycer
18	293.5	13.4	548	2 S37969	probable triacygl
19	259.5	11.8	413	2 T43170	probable triacygl
20	254	11.6	538	2 S64842	probable membrane
21	237	10.8	573	2 S84754	probable membrane
22	212	9.7	473	2 D86318	protein R15H18.6 [
23	170.5	7.8	509	2 G96766	protein lipase F2P
24	165	7.5	336	2 F83425	probable esterase/
25	147	6.7	460	2 G96764	unknown protein F2
26	121.5	5.5	456	2 E89391	hypothetical prote
27	110.5	5.0	999	2 B70501	hypothetical prote
28	108.5	4.9	1002	2 D87077	probable long-chain
29	107.5	4.9	841	2 A90669	probable enzyme [i

30	107.5	4.9	841	2 D85519	probable enzyme ya
31	107.5	4.9	841	2 C64755	yagX protein - Bsc
32	100	4.6	1216	2 F88473	protein F40H6.5 [i
33	99.5	4.5	798	2 F96714	probable protease
34	99	4.5	323	2 S1927	lipase A precursor
35	98	4.5	465	1 JC1318	triacylglycerol li
36	97.5	4.4	569	2 S75169	urease (EC 3.5.1.5
37	95.5	4.4	461	2 S21223	triacylglycerol li
38	95.5	4.4	467	1 LIDG	triacylglycerol li
39	95	4.3	673	2 T24347	hypothetical prote
40	95	4.3	1648	2 S61654	probable membrane
41	95	4.3	2100	2 T03223	probable polyketid
42	94.5	4.3	835	2 T35763	probable aminopapt
43	94	4.3	519	2 A50726	probable membrane
44	94	4.3	554	2 A47503	epoxide hydrolase
45	94	4.3	569	2 A97648	urease alpha chain

ALIGNMENTS

RESULT 1

E84526  
Probable lysosomal acid lipase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-feb-2001  
C:Accession: E84526  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: R84420; MUID:20083487; PMID:10617137  
A:Accession: E84526  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-344 <STO>  
A:Cross-references: GB:AE002093; NID:q4585908; PIDN:AAD25569.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: Atg15230  
A:Map position: 2

Query Match	48.1%	Score 1055;	DB 2;	Length 344;
Best Local Similarity	54.5%;	Pred. No. 4.2e-82;		
Matches 198;	Conservative 42;	Mismatches 71;	Indels 52;	Gaps 2;
QY	40	GGSGGLCDQLLLPLGYPCTEHNVTGDKGFLSLQHPHGKNAADSTGPPVFLQHLGFLQ	99	
Db	26	GSPVNSLCADLIHPANSCTEHSIQTKDGYILALQVAVSLGPRLL--OSGPPVLLQHLGFLM	83	
QY	100	GGDTWFINSAEQSLGYILADNGFDVWGNVGRTRMSKGHSTFVHDKLFDWDSWQELAEY	159	
Db	84	AGDVWFLNSPKESLCGFIADHGFDVWGNVGRTRYSGHVTLSDDTKDFWDSWQDLAMY	143	
QY	160	DLAMLGLVYVTVTSQKILYVGHSGQITMGLAALTWPEIVKMKISSAALLCPSYLDHVSAS	219	
Db	144	DLAEMIQYLSISNSKIFLVGHSGQITMSFAALTQPHVAEMVEAAALLCPSYLDHVTAP	203	
QY	220	FVLRAVAMHLDQMLVTMGTHQLNFRSDMGVQIVDSLDCGEHVDCNNLLSAITGENCCFNT	279	
Db	204	LVERVMFVHLDQ-----	215	
QY	280	SRIDYLYLEYPHPSPSTKNLHLFQMRKTFYAKYDYGLLGNLRRYCHLRPPAFDLSIPE	339	
Db	216	--IEYLDYEPHPSSVKNIHLFQMRKTFYAKYDYGFKNLRTYGLSKPPEFILSHIPA	273	
QY	340	SLPTWMYGGGLDALADVTDVORTIRELGSTPELLYIGDYGHIDFVMSVKAKDDVYVDLIR	399	
Db	274	SLPVMWVGCTDGLADVDVDEHTLAELFSSPELLYLEDYGHIDFVIGSSAKEDVYKHMIO	333	
QY	400	FLR 402		
Db	334	FFR 336		

## RESULT 2

T33198

hypothetical protein ZK6.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T33198

R:Wu, X.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid ZK6.

A:Reference number: Z21301

A:Accession: T33198

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-403 &lt;WUX&gt;

A:Cross-references: EMBL:AF067942; PIDN:AA017694.1; GSPDB:GN00023; CESP:ZK6.7

A:Experimental source: strain Bristol N2; clone ZK6

C:Genetics:

A:Gene: CESP:ZK6.7

A:Map position: 5

A:Introns: 29/3; 63/1; 219/3; 319/3; 365/3

C:Superfamily: triacylglycerol lipase, lingual

Query Match 25.4%; Score 557.5; DB 2; Length 403;  
Best Local Similarity 32.4%; Pred. No. 1.3e-39;  
Matches 124; Conservative 65; Mismatches 153; Indels 41; Gaps 11;

Qy 49 QLLPLGYPCTEHNVTGKDFLLSLQHIPHGKNAADSTG--PPVFLQHLFGQGDWTFI 106

Db 31 QIIERWGPAMYYTATDDGVILEMHRIPFGKTNVTPNGKRPVVFQWHLGCASSDMMV 90

Qy 107 NSAEQSLGYILLADNGFDWVIGNVRGRWSKSHSTFSVHDKLFDWMSWOELAEYDILLAMLG 166

Db 91 NLPOGAGFLADAGFDVWLNMRGNTYSMKHDKLPSHSAFWDSDWDEMATYDLNAMIN 150

Qy 167 YVYTVT-OSKILYVGHSGTGTGMLAALTPME--IVKMISAAALCPISYLDHVSAA--SFV 221

Db 151 HVLEVTGQDSVYVGHSGTGTGMLAALTPME--IVKMISAAALCPISYLDHVSAA--SFV 210

Qy 222 LEAVAMHLDQMLVYVGHSGTGTGMLAALTPME--IVKMISAAALCPISYLDHVSAA--SFV 274

Db 211 ANYFSLEFDGWFDFGFEFLPNWAMKLAAR---DICGGGLKVEADLCNDVLFLLIAGES 267

Qy 275 CFNFSRIDYILEYEPHSPSTKNLHHLFQMIKRGTFYKDYGLNLRYYGHLRPPAPDL 334

Db 268 DOWNTRVPVYATHDPAGTSTQNIYVHWQVYHGGVPAYDAGTKTNKKYGOANPEYDF 327

Qy 335 SSIPESLPIMWYGGGLDALADVTQVIRTELGSSTPELL-----YIGDYGHI 381

Db 328 TAI-KGTIYLYWSADWLADTPDV-----PDYLLRLNPAIVAQNHLDPYNNHL 376

Qy 382 DPMVSVKAKDDVYVDLIRFLREN 404

Db 377 DPTWGLRAPDDIYRPAIKLCTDD 399

## RESULT 3

JC4017

triacylglycerol lipase (EC 3.1.1.3) PGE precursor - bovine

N:Alternate names: pregastric esterase

C:Species: Bos primigenius taurus (cattle)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: JC4017; S64678

R:Timmermans, M.Y.J.; Teuchy, H.; Kupers, L.P.M.

Gene 147, 259-262, 1994

A:Title: The cDNA sequence encoding bovine pregastric esterase.

A:Reference number: JC4017; MUID:95011625; PMID:7926811

A:Accession: JC4017

A:Molecule type: mRNA

A:Residues: 1-397 &lt;TIM&gt;

A:Cross-references: GB:L26319; NID:g600756; PIDN:AAA57037.1; PID:g600757

A:Experimental source: tongue

Query Match 25.0%; Score 549; DB 2; Length 405;

Best Local Similarity 31.9%; Pred. No. 7.1e-39;

R:Timmermans, M.Y.J.; Reekmans, G.; Teuchy, H.J.H.; Kupers, L.P.M.

Biochem. J. 314, 931-936, 1996

A:Title: Inhibition studies on calf pregastric esterase: the enzyme has no functional

A:Reference number: S64678; MUID:96177869; PMID:8615791

A:Accession: S64678

A:Molecule type: protein

A:Residues: 24-40; 248-253 &lt;TIW&gt;

C:Comment: Pregastric esterase is a major fat-digesting enzyme.

C:Genetics:

A:Gene: pge

C:Superfamily: triacylglycerol lipase, lingual

C:Keywords: blocked amino end; carboxylic ester hydrolase; glycoprotein

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-397/Product: pregastric esterase #status predicted &lt;MAT&gt;

F:33,270,326/Binding site: carboxylate (Asn) (covalent) #status predicted

F:171/Active site: Ser #status predicted

Query Match 25.2%; Score 553; DB 1; Length 397;

Best Local Similarity 34.9%; Pred. No. 3.2e-39;

Matches 130; Conservative 67; Mismatches 146; Indels 30; Gaps 13;

Qy 49 QLLPLGYPCTEHNVTGKDFLLSLQHIPHGKNAAD-STGPPVFLQHLFGQGDWTFIN 107

Db 36 QMISYWGYPSEMHKVITADGVILQVYRIPHGKNNANHLGQRPVYVFLQHLGSGSATNWSN 95

Qy 108 SAEQSLGYILLADNGFDWVIGNVRGRWSKSHSTFSVHDKLFDWMSWOELAEYDILLAMLG 167

Db 96 LPKNSLGLFLLADAGYDVLWNGSRGTWQAEHLHYSPDSPFPAFSDMAEYDLPTIDF 155

Qy 168 YVYTVT-OSKILYVGHSGTGTGMLAALTPMEIVKMSAALLCPISYLDHVSASFVLRV 225

Db 156 ILRRTGQKLLHYVGHSGTGTGTFIAFSTPLAEKIKFYALAPVATVTKYKSLFNKAL 215

Qy 226 AMHL-----DQMLVTMGTHOLNFRSDMGVQIVDSLDCDGEHVD--CNNLLSAITG-ENC 275

Db 216 IPHFLFKIIFGDKMFPY--HTF-LEQFLGVE---MCSRETLDVCKNALFAITGVDNK 267

Qy 276 CFNFSRIDYILEYEPHSPSTKNLHHLFQMIKRGTFYKDYGL-LGNLRRYGHLPAPDL 334

Db 268 NFNSRLDVIYAHNPAGTSVQNTLHWQAVKSGKGFQAFDMGAPYQNLHMYHQPPTIYNL 327

Qy 335 SSIPESLPIMWYGGGLDALADVTQVIRTELGSSTPELLY---IGDYCHIDFVMSVAKD 391

Db 328 TAMVPAVMSADN--DLLADPDQVDFLLSKLSN---LIYHKEIPNPNHLDIFWAMDAPQ 382

Qy 392 DVYVDLIRFLREN 404

Db 383 EYNEIVSLMAED 395

## RESULT 4

T22675

hypothetical protein F54F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T22675

R:Percy, C.; Lloyd, C.

submitted to the EMBL Data Library, September 1996

A:Reference number: Z19598

A:Accession: T22675

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-405 &lt;WIL&gt;

A:Cross-references: EMBL:Z79696; PIDN:CAB01973.1; GSPDB:GN00023; CESP:F54F3.3

A:Experimental source: clone F54F3

C:Genetics:

A:Gene: CESP:F54F3.3

A:Map position: 5

A:Introns: 31/3; 65/1; 274/3; 366/3

C:Superfamily: triacylglycerol lipase, lingual

Query Match 25.0%; Score 549; DB 2; Length 405;

Best Local Similarity 31.9%; Pred. No. 7.1e-39;







C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C;Accession: G89074  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999; and  
A;Accession: G89074  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-411 <STO>  
A;Cross-references: GB:chr\_V; PIDN:AAAC48051.1; PID:g1658354; GSPDB:GN00023; CESP:K04A8.5  
C;Genetics:  
A;Gene: K04A8.5  
A;Map position: 5  
C;Superfamily: triacylglycerol lipase, lingual

Query Match 21.3%; Score 466.5; DB 2; Length 411;  
Best Local Similarity 32.4%; Pred. No. 7.8e-32;  
Matches 118; Conservative 62; Mismatches 163; Indels 21; Gaps 11;

Qy 49 QLLPLPGPCTEHNVTGKDFLLSLQIHPGCK----NKAADSTGPPVFLQHLFGQGGTW 104  
Db ELIKSWGYSVEIYNTTKDGFLELHRIPYGREVPTSDVNNRPVIFLQHLGFLCSSFDW 97  
Qy 105 FINSAEQSLGYLADNGFDVWGNVGRNWSKGHSTFSV-HDKLFDWMSQWELAEYDLA 163  
Db VANSPHQAGVFADAGFDVWGNFRGNTYSRKHVSLNPKDPKFDWMSQWDSIYDLPA 157  
Qy 164 MLGVYVTVT-OSKILYVGHSGGTGIMLAAL-TMPEIVKMISSAALLCPISYLDHVSASFV 221  
Db MIGRALEISGQESLYTGFSLGTLTMEAKLSTDPKFSKIKKYPALAPIGSIKHAHVFL 217  
Qy 222 L--RAVAMHLQMLVTMGIHOLNFRSDMGVQIVDSLCL--DGEHVDCNNLSAITS--E 273  
Db FLGRHFGKDYEEYVKKHSGDELFSSLLKVIKYTCGLFTLEEFCSDFITLLFIGTANE 277  
Qy 274 NCCFNTSRIDYLEYEPHPSSTKNLHLLFQMRKGTAKYDGLLGNLRRYGHLPAPFD 333  
Db N--WQTRIPVLAHTPAGSSSNVMAHLDMQFSGYGVYPTFDMGEKNLKAQYQKLPQYN 335  
Qy 334 LSSIPESLPWMYGGGLDALADVDVQRTI-RELGS--TPPELLYIGDYGHIDFVMSKAK 390  
Db FTGIAD-VPIVFLWDDDLSTKQLEETLFAQLNSQVVGFSRIENYNHLHFIMGTTAA 394  
Qy 391 DDVY 394  
Db SQVY 398

RESULT 13  
JT0949  
egg-specific protein - silkworm  
C;Species: Bombyx mori (silkworm)  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 26-Feb-1998  
C;Accession: JT0949; A28527  
R;Sato, Y.; Yamashita, O.  
submitted to JIPID, September 1991  
A;Reference number: JT0949  
A;Accession: JT0949  
A;Molecule type: DNA  
A;Residues: 1-559 <SAT>  
A;Experimental source: larva  
A;Note: This protein is a homotrimer  
R;Indrasith, L.S.; Sasaki, T.; Yamashita, O.  
J. Biol. Chem. 263, 1045-1051, 1988  
A;Title: A unique protease responsible for selective degradation of a yolk protein in Bombyx mori  
A;Reference number: A28527; MUID:88087166; PMID:3275655  
A;Accession: A28527  
A;Molecule type: protein  
A;Residues: 19-40;133-144, 'T', 146-152;229-248 <IND>  
C;Genetics:

A;Map position: 19  
C;Keywords: egg Yolk; homotrimer  
F;132-133/Cleavage site: Lys-Asn (egg 30.5K cysteine proteinase) #status experimental  
F;228-229/Cleavage site: Arg-Asp (egg 30.5K cysteine proteinase) #status experimental

Query Match 19.0%; Score 416.5; DB 2; Length 559;  
Best Local Similarity 31.9%; Pred. No. 2.2e-27;  
Matches 122; Conservative 64; Mismatches 152; Indels 45; Gaps 17;

Qy 49 QLLPLPGPCTEHNVTGKDFLLSLQIHPGCKNKAADSTGPPVFLQHLFGQGGTWFIN 108  
Db ELIKKHQYPVEEHTVATDDGYHLTVLRIPPTHQTRDDKKPVALMHGLLSADDWLLMG 253  
Qy 109 AEQSLGYLADNGFDVWGNVGRNWSKGHSTFSVHDKL--FWDMSQWELAEYDLA 166  
Db PSKSLAYMLSDAGYDVLGNVGRNYSRSH--VSKHPALNDFNKFSDNDEIALHDLPAID 311  
Qy 167 YVTVT-OSKILYVGHSGGTGIMLAALT-MPEIVKMISSAALLCPISYLDHVSASFV 224  
Db HVLDISQGERLHYGHSGGATTFALMSEQPSYNEKIVSMHALSPIVMYV-RSPLFRM 370  
Qy 225 VA-----MHLDOM---LVTMGIHOLN-----FRSDMGVQIVDSLCLDGEHVD 267  
Db IAPTSKPYQYIH-DQVGHGAFEPGKHLIETFGNACREKLCR-----HV-CNNLN 419  
Qy 268 SAITGENCEPNTSR--IDYYLEYEPHPSSTKNLHLLFQMRKGTAKYDGLLGNLRRY 325  
Db YVSGIN-VYNQADIVPVVMAHLPACTSARVMKQYQNVASHDFRKYNYGAETNMKVYG 478  
Qy 326 HLPPAPDLSIPESLPWMYGGGLDA-LADVTDVQRTIRELGSSTPELLYIGDYGH--I 381  
Db ASEPPSYDLKSV--SAPVNL-YHSHDAWLHPKDVEXLQENLPVKOSFEVPEQOHTDL 535  
Qy 382 DFVMSVAKAKDDVYVDLIRFLREN 404  
Db DFQFSKKAPDVTYOKLMENQNN 558

RESULT 14  
T31611  
hypothetical protein Y50E8A.g - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T31611  
R;Steward, C.  
submitted to the EMBL Data Library, September 1999  
A;Reference number: Z21047  
A;Accession: T31611  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1585 <WIL>  
A;Cross-references: EMBL:AL117200; NID:ei549770; PIDN:CAB55050.1; CESP:Y50E8A.g  
A;Experimental source: clone Y50E8A  
C;Genetics:  
A;Gene: CESP:Y50E8A.g  
A;Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match 17.9%; Score 392.5; DB 2; Length 1585;  
Best Local Similarity 27.5%; Pred. No. 9.6e-25;  
Matches 114; Conservative 68; Mismatches 188; Indels 45; Gaps 13;

Qy 11 ALPILLVWSCGRIVGASPAALARRVGSGLCDQLLLPLGYPCTEHNVTGKDFL 70  
Db AVNFFIVFAFGKEIETLDAADTISH-----YGYTVKHYVTDDGYT 48  
Qy 71 LSLQIHPGCKNKA--DSTGPPVFLQHLFGQGGTWFINSAEQSLGYLADNGFDVWIG 127  
Db VOLQIRIPVGRDRLSILGCKSRPVVFFMHGLFGSSVHFLLNLPSSAAVIFADAGFDW 108  
Qy 128 NVGRTRWSKGHSTFSVHDKLFDWMSQWELAEYDLA 186  
Db NIRTGYGLNHTSFSTNGVNFNFSLEYHSHYDLRQQLYVLDYTRHESFLSYVGHSG 168

```
QY 187 MGLAALTMPEIV-----KMISAAALLCPISYLDHVSASFVLRVAVAMHLDQMLVTWG 237
Db 169 VMFARLAEDVTWQSKIRVFFALPTAGFMKPLPFTLLLENYLQALIQFALDG---KFG 225
QY 238 IHOENFRSDMGVQIVDSLCDGSHVD--CNLLSAITGENCC--FNTSRIDYVLEYEPHPS 293
Db 226 ILPVEIPRAIASKFAD-FCSSKFFTFELCSAGFKVAAGIETLGOVNDSDRIPIILSHFPSAT 284
QY 294 STKNLHLHFMIRKGTFAKYDYGLGNLRRYRPPAFDLSSIPESLPIMWGYGGLDAL 353
Db 285 STLNLHWMQIFKYHELRLDLGTARNLIAYGQKAPRLEIGNIIAQITLY--FSKDDQI 342
QY 354 ADVTDVQRTI-RELGSTPELLYIGD---YGHIDFVMSVKAKDDVYVDLI-RELRE 403
Db 343 TDEVDVREIIMKMG--FGLIESYLDHFTFDFILGLRATDEVYKPIVYRIYKE 395

RESULT 15
T39540
triglyceride lipase-cholesterol esterase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T39540
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A:Reference number: 221862
A:Accession: T39540
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-443 <WOO>
A:Cross-references: EMBL:AL021748; PIDN:CAA16863.1; GSPDB:GN00067; SPDB:SPBC16A3.12c
A:Experimental source: strain 972h-; cosmid c16A3
C:Genetics:
A:Gene: SPDB:SPBC16A3.12c
A:Map position: 2
A:Introns: 11/2; 152/3; 282/1; 398/3
C:Superfamily: triacylglycerol lipase, lingual

Query Match 16.2%; Score 356; DB 2; Length 443;
Best Local Similarity 28.3%; Pred. NO. 2.3e-22;
Matches 108; Conservative 75; Mismatches 149; Indels 50; Gaps 16;

QY 55 GYPCTEHNVEYKDGFLLSLOHPIHGKNAADSTGPPVFLHGLFOGGDTWF-INSABOSL 113
Db 83_GYRVEEHLVRTQDNFILCLHRIHPKQ--SQHKREVVYCHHGLMTNSELWAVNESERSL 140
QY 114 GYILADNGFDVWIGNVGRTRNSKGHSTFSVHDKLFWDKSWQELAEYDILLAMLGYYVTY- 172
Db 141 PVFLIESGTDVWGLNNGRNKYSRKHIITYKPRDEEFNFSLDDMAMFDIPDTVDYILRETG 200
QY 173 QSKILYVGHSGQTIMGLAALTM-PEIVKMISAAALCP-----ISYLDHVSASF 220
Db 201 REKLVNIGFSQCTAQAMAALSINPDNDKNVIFIGLAYAPKGFSNFYVDYIVKVNPKI 260
QY 221 VLRVAMHLDQMLVTMGTHQLNFRSDMGVQIVD-SLCDGEHVDCNNLLSAITGENCCPNT 279
Db 261 MYHLEGRRC--LLPSVTFWQNICYPPIEFKIVDVS-----KILFNWDLNLSLNQ 309
QY 280 SRIDYVLEYEPHPSSTKNLHLFOMIRKGTFAKYDYGLGNLRRYG--HLRPPAFDLSSI 337
Db 310 KLCGYAHLYS--FSSVKSVMHVLIIKIKTQLYDQDD--MALLAGYGSRRHYQVPLEPTNNI 366
QY 338 PESLPIMWGYGGLDALADVTQVORTIRELGSTPELLYIGDYGHIDFVMSVKAKDDVY--- 394
Db 367 --KCPMLILWGGKDTLINN-EVMRT--ALPPHAKVSIATYEHLDLFWGQDVKEEVPV 421
QY 395 VDLIR-----FLRENGWH 407
Db 422 IDALKHSHSLGKAKHEVKQNGEH 443
```

Search completed: December 30, 2002, 20:47:19  
Job time : 53 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2002, 19:06:36 ; Search time 38 Seconds  
(without alignments)  
447.508 Million cell updates/sec

Title: US-09-699-652A-14  
Perfect score: 2194  
Sequence: 1 MAMAGHAPGALPLILLVVS.....DDVYVDLIRFLRENGWHNSY 410

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	553	25.2	397	1	LIPG_BOVIN
2	546	24.9	398	1	LIPG_HUMAN
3	539.5	24.6	395	1	LIPG_RAT
4	529	24.1	398	1	LIPG_CANFA
5	515.5	23.5	399	1	LICH_HUMAN
6	513.5	23.4	394	1	LIP3_DROME
7	503	22.9	397	1	LICH_MOUSE
8	486	22.2	397	1	LICH_RAT
9	427	19.5	439	1	LIP1_DROME
10	293.5	13.4	548	1	TGL1_YEAST
11	107.5	4.9	841	1	YAGX_ECOLI
12	100	4.6	1216	1	YKX5_CAEEL
13	97.5	4.4	569	1	URE1_SYNV3
14	95.5	4.4	461	1	LIPP_HORSE
15	94	4.3	465	1	LIPP_RABIT
16	94	4.3	554	1	HVES_RAT
17	93.5	4.3	467	1	LIP1_CANFA
18	93.5	4.3	1167	1	CIJA_BACTU
19	93.5	4.3	1680	1	C05_MOUSE
20	92	4.2	239	1	SFSA_AGR75
21	91	4.1	465	1	LIPP_CAVPO
22	91	4.1	569	1	URE1_BACSU
23	90.5	4.1	465	1	LIPP_HUMAN
24	90	4.1	1676	1	C05_HUMAN
25	89.5	4.1	445	1	SCRC_RABIT
26	88.5	4.0	518	1	YONE_ECOLI
27	88.5	4.0	669	1	GALC_CANFA
28	88.5	4.0	1876	1	PPSA_MYCTU
29	88	4.0	457	1	LIPP_MYOCO
30	88	4.0	972	1	KFMS_HUMAN
31	86.5	3.9	281	1	YN07_MYCTU
32	86.5	3.9	307	1	PTRB_STRPN
33	86	3.9	519	1	DPYS_MOUSE

ALIGNMENTS

RESULT 1

ID	LIPG_BOVIN	STANDARD;	PRT;	397 AA.
AC	Q29458;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Triacylglycerol lipase, pregastric precursor (EC 3.1.1.3) (Pregastric			
DE	lipase) (GL) (Pregastric esterase) (PGE).			
GN	LIPF.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Torque serous gland;			
RX	MEDLINE=95011625; PubMed=7926811;			
RA	Timmermans M.Y.J., Kupers L.P., Teuchy H.;			
RT	"The cDNA sequence encoding bovine pregastric esterase.";			
RL	Gene 147:259-262(1994).			
CC	-!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a			
CC	fatty acid anion.			
CC	-!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: L26319; AAA57037.1;			
DR	InterPro: IPR000073; Abhydrolase.			
DR	InterPro: IPR000734; Lipase.			
DR	InterPro: IPR000379; Ser_estrs_site.			
DR	Pfam: PF00561; abhydrolase; 1.			
DR	PROSITE: PS00120; LIPASE_SER; 1.			
KW	Hydrolase; Lipid degradation; Glycoprotein; Signal.			
FT	SIGNAL 1 19			
FT	CHAIN 20 397			
FT	ACT_SITE 171 171			
FT	ACT_SITE 371 371			
FT	CARBOHYD 33 33			
FT	CARBOHYD 270 270			
FT	CARBOHYD 326 326			
FT	SEQUENCE 397 AA; 45231 MW; F68977DED585EE36 CRC64;			
SQ				
Query Match	25.2%;	Score 553;	DB 1;	Length 397;
Best Local Similarity	34.9%;	Pred No. 3.1e-40;		
Matches 130;	Conservative 67;	Mismatches 146;	Indels 30;	Gaps 13;
QY 49	QLLLPLGYCTEHNVTKDFLLSLQHPGKNKAAD-STGPPVFLQHGLFGOGDTWFIN 107			

Db 36 QMISYWGYPSEMHKVIITADGYILOVYRIPHGKNNANHLGQRPVVFLQHLGSGATNNISN 95  
Qy 108 SAEQSLGYILLADNGFDWIGNVGRTRWSKGHSTESVHDKLFDWMSWOELAEYDILLAMGY 167  
Db 96 LPKNSLGLFLADAGYDVLGNSRGTWAQEHLYSPSPFEWAFSEFDEMAEYDLPSTIDF 155  
Qy 168 VYTVT-QSKILYVGHSGQTGLAAL-TMPEIVKMISAAALLCPISYLDHVSASFVLRV 225  
Db 156 ILRRTGQKLLHYVGHSGQTGTGFTAFSTPTLAEKIKFVALAPVATVKYTKSLFNKLAL 215  
Qy 226 AMHL-----DQMLVTMGTHQLNFRSDMGVQIVDSLCDGEHVD--CNNLLSAITG-ENC 275  
Db 216 IPHFLFKIIFDKMFP---HTF-LEQFLGVE----MCSRETLVLCCKNALFAITGVDNK 267  
Qy 276 CFNFSRIDYLYEYPHPSSTKNLHLHLMQIRKGFYAKYDYL-LGNLRRYGHLPAPFDL 334  
Db 268 NFNSRLDVIYIAHPAGTSVQNTLHWQAVKSGKFQFADGAPQNLHMYHQPPTPIYNL 327  
Qy 335 SSIPESLPIMWYGGGLDALADVTQVIRELGSTPELly---IGDYCHIDFVMSVRAKD 391  
Db 328 TAMNVPIAVWSADN-DLLADPQDVDFLLSKLSN--LIYHKEIPNYNHLDFIAMDAPQ 382  
Qy 392 DYYVDLIRFLREN 404  
Db 383-EVYNEIVSLMAED 395  
RESULT 2  
ID LIPG-HUMAN STANDARD; PRT; 398 AA.  
AC P07098;  
DC 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Triacylglycerol lipase, gastric precursor (EC 3.1.1.3) (Gastric  
lipase) (GL).  
GN LIPG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87299724; PubMed=3304425;  
RA Bodmer M.W., Angel S., Yarranton G.T., Harris T.J.R., Lyons A.,  
RA King D.J., Pieroni G., Riviere C., Verger R., Lowe P.A.;  
RT "Molecular cloning of a human gastric lipase and expression of the  
RT enzyme in yeast."  
RL Biochim. Biophys. Acta 909:237-244(1987).  
RN [2]  
RP SEQUENCE OF 20-45.  
RX MEDLINE=89325292; PubMed=2753032;  
RA Bernbaeck S., Blaeckberg L.;  
RT "Human gastric lipase. The N-terminal tetrapeptide is essential for  
RT lipid binding and lipase activity."  
RL Eur. J. Biochem. 182:495-499(1989).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE=99287897; PubMed=10358049;  
RA Roussel A., Canaan S., Egloff M.P., Riviere M., Dupuis L., Verger R.,  
RA Cambillau C.;  
RT "Crystal structure of human gastric lipase and model of lysosomal  
RT acid lipase, two lipolytic enzymes of medical interest."  
RL J. Biol. Chem. 274:16995-17002(1999).  
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X05997; CAA29413.1; -  
DR EMBL; X05997; CAA29414.1; ALT\_INIT.  
DR EMBL; A01046; CAA00125.1; -  
DR EMBL; A12714; CAA01053.1; -  
DR PIR; S04942; S04942.  
DR PIR; S07145; S07145.  
DR PDB; 1HLG; 15-MAR-00.  
DR Genes; HGNC:6622; LIPF.  
DR MIM; 601980; -  
DR InterPro; IPR000073; Abhydrolase.  
DR InterPro; IPR000734; Lipase.  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR Pfam; PF00561; abhydrolase; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
KW Hydrolase; Lipid degradation; Glycoprotein; Signal; Polymorphism;  
KW 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 398 TRIACYLGLYCEROL LIPASE, GASTRIC.  
FT ACT\_SITE 172 172 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 343 343 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 372 372 CHARGE RELAY SYSTEM.  
FT DISULFID 246 255  
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .).  
FT VARIANT 161 161 /FTID=VAR\_011947.  
SQ SEQUENCE 398 AA; 45237 MW; CD3EE1621C014F0F CRC64;  
Query Match 24.98; Score 546; DB 1; Length 398;  
Best Local Similarity 33.98; Pred. No. 1.2e-39;  
Matches 130; Conservative 70; Mismatches 132; Indels 52; Gaps 13;  
Qy 49 QLLPLPGYPCPEHNVETKDGFLLSLQHIPHGKNAADSTG--PPVFLQHLGFGQGDWTFI 106  
Db 37 QMIYWGYPNEEYEVWTEGYILEVNRIPYCK-KNSGNTGQRPVVFLQHLGSLASATNWS 95  
Qy 107 NSAEQSLGYILLADNGFDWIGNVGRTRWSKGHSTESVHDKLFDWMSWOELAEYDILLAMG 166  
Db 96 NLPNNSLAFLADAGYDVLGNSRGTWAQEHLYSPSPFEWAFSEFDEMAEYDLPATID 155  
Qy 167 VYTVT-QSKILYVGHSGQTGLAAL-TMPEIVKMISAAALLCPISYLDHVSASF-VLR 223  
Db 156 FIVKTKGQKLLHYVGHSGQTGTGFTAFSTNPFLAKRIKTFYALAPVATVKYTKSLINKLR 215  
Qy 224 AVAMHL-----DQMLVTMGTHQLNFRSDMGVQIVDSLCDGEHVD--CNN 265  
Db 216 FVPSQLFKFIQDKTFYPHNFDFQFLAT-----EYCSREMLNLLCSN 257  
Qy 266 LLSAITG-ENCCFNSTRIDYLYEYPHPSSTKNLHLHLMQIRKGFYAKYDYL-LGNLRR 323  
Db 258 ALFIICGFDKSNFNTSRDLYVLSHNPAGTSVQNMFHMTQAVKSGKFQAYDWGSPVQNRMH 317  
Qy 324 YGHLRPPAFDLSLSPESLPIMWYGGGLDALADVTQVIRELGSTPELly---IGDYGH 380  
Db 318 YDQSQPPYINVTAMNVPJAVW--NGGKDLADPQDVQVGLLLPKL----PNLIYHKEIPFTYNNH 372  
Qy 381 IDFVMSVRAKDDYVVDLIRFLREN 404  
Db 373 LDFIAMDAPQEVYNDIVSMISED 396  
RESULT 3  
ID LIPG-RAT STANDARD; PRT; 395 AA.  
AC P04634;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)





```
FT VARIANT 16 16 P -> T.
FT VARIANT 129 129 H -> P (IN CESD).
FT VARIANT 129 129 H -> P (IN CESD).
FT VARIANT 129 129 H -> R (IN CESD).
FT VARIANT 200 200 L -> P (IN CESD AND WD).
FT VARIANT 23 23 G -> R (IN REF. 3).
FT VARIANT 29 29 V -> L (IN REF. 4).
SQ SEQUENCE 399 AA: 45415 MW: 55F27391306B609A CRC64;

Query Match 23.5%; Score 515.5; DB 1; Length 399;
Best Local Similarity 32.3%; Pred. No. 5.4e-37;
Matches 131; Conservative 59; Mismatches 143; Indels 73; Gaps 12;

QY 42 GSGGL-----CDQLLLPLGVPCTEHNVEKDGFLLSLQHTPHG-KNKAADSTGP 89
DB 21 GSGGLTAVDPETNNVSEIISYWGFPSEYIYVEDGVLCLNRIPIHGRKKNHSDKGP 80
QY 90 PVFLQHLGFGQGDWFIINSAEQSLGIYILADNGFDVWIGNVGRTRWSKGHSTFVHDKLFW 149
DB 81 VVFLQHLGLLADSSNVNTLANSLSGIFLADAGFDVWGMGRGNTWSRKHKTLSVSDQDEF 140
QY 150 DMSWELAEYDLMLGYYVTVT-OSKILYVGHSGOTINGLAALT-MPIVIVKMSAALL 207
DB 141 AFSYDEMAKYDLPASINFTLNKTGQEQVYVGHSGQTTIGTFAFSAIPELAKRIRMFAL 200
QY 208 CPISYLDHVSASFVLRVAVM-----HLDQMLVTWGIHOLNFRSDMGVQIVDSLCD---- 257
DB 201 GPVA-----SVAFCFSMAKLRGLPDHL-----IKDLFGDKFEL 234
QY 258 -----GEHVDCCNNLSAITGNCFC-----NTSRIDYILEYEPHPSSTKNL 298
DB 235 POSAEKWLGTHTVCHVILKELCG-NLCELLCGFERNLNNSRVDVYTHSPAGTSVQNM 293
QY 299 HHLFQMRKGTFAKYDYLGLG-NLRRYGHLPAPFDLSSIPESLPIMWGMGLDALADVT 357
DB 294 LHWQAVKFKQFAEDWGSNAKNEYHNSQYPTNVKMDLPTAVWS--GGHDWLADVY 351
QY 358 DVQRTIRELGSPELLYIGDYGHIDFVMSVAKDDVYVDLIRLE 403
DB 352 DVNILLTQITNLVTHFESPEWEHLOFVGLDAPWRLNKKIINLMRK 397

RESULT 6
ID LIP3_DROME
AC 046108; STANDARD; PRT; 394 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Lipase 3 precursor (EC 3.1.1.-).
GN LIP3 OR C68823.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=9827315; PubMed=9566193;
RA Pistillo D., Manzi A., Tino A., Pilo Boyl P., Graziani F., Malva C.;
RT "The Drosophila melanogaster lipase homologs: a gene family with
RT tissue and developmental specific expression.";
RL J. Mol. Biol. 276:877-885(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
```

```
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabriellian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- TISSUE SPECIFICITY: FAT BODY.
CC -!- DEVELOPMENTAL STAGE: ONLY AT LARVAL STAGES.
CC -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.). ALSO SIMILAR TO
CC LEPIDOPTERAN EGG-SPECIFIC AND YOLK PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y14367; CAA74737.1;
CC EMBL: AE003699; AAF54935.1;
CC FlyBase: FBgn0023495; Lip3.
CC InterPro: IPR000073; Abhydrolase.
CC InterPro: IPR000734; Lipase.
CC InterPro: IPR000379; Ser_estr_site.
CC Pfam: PF00561; abhydrolase_1.
CC PROSITE: PS00120; LIPASE_SER; 1.
KW Hydrolase; Lipid degradation; Signal; Glycoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 394 LIPASE 3.
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 369 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 131 131 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 394 AA: 44901 MW: A718D1D743673802 CRC64;

Query Match 23.4%; Score 513.5; DB 1; Length 394;
Best Local Similarity 31.4%; Pred. No. 7.9e-37;
Matches 128; Conservative 78; Mismatches 162; Indels 39; Gaps 14;

QY 10 GALPILLLVWSCCGRIVSGASPAALRRVSGSGGLCDQLLLPLGYPCTEHNVEKDFG 69
DB 4 GALKVTILLVG-LGLVLGSRPISD-----CGERIEDDGYPMERHEVVTSDNY 50
```



RX MEDLINE=96129534; PubMed=8576647;  
RA Nakagawa H., Matsubara S., Kuriyama M., Yoshidome H., Fujiyama J.,  
RA Yoshida H., Osame M.;  
RT "Cloning of rat lysosomal acid lipase cDNA and identification of the  
RT mutation in the rat model of Wolman's disease.";  
RL J. Lipid Res. 36:2212-2218(1995).  
CC -!- FUNCTION: CRUCIAL FOR THE INTRACELLULAR HYDROLYSIS OF CHOLESTERYL  
CC ESTERS AND TRIGLYCERIDES THAT HAVE BEEN INTERNALIZED VIA RECEPTOR-  
CC MEDIATED ENDOCYTOSIS OF LIPOPROTEIN PARTICLES. IMPORTANT IN  
CC MEDIATING THE EFFECT OF LDL (LOW DENSITY LIPOPROTEIN) UPTAKE ON  
CC SUPPRESSION OF HYDROXYMETHYLGLUTARYL-COA REDUCTASE AND ACTIVATION  
CC OF ENDOGENOUS CELLULAR CHOLESTERYL ESTER FORMATION (By  
CC similarity).  
CC -!- CATALYTIC ACTIVITY: A sterol ester + H(2)O = a sterol + a fatty  
CC acid.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S81497; AAB36043.2; -;  
DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR000734; Lipase.  
DR InterPro: IPR000379; Ser\_estrs\_site.  
DR Pfam: PF00561; abhydrolase; 1.  
DR PROSITE; PS00120; LIPASE\_SER; FALSE\_NEG.  
KW Hydrolase; Lipid degradation; Glycoprotein; Signal; Lysosome.  
FT SIGNAL 1 19  
FT CHAIN 20 397  
FT LYSOSOMAL ACID LIPASE/CHOLESTERYL ESTER  
FT HYDROLASE.  
FT ACT\_SITE 172 172  
FT ACT\_SITE 372 372  
FT CARBOHYD 34 34  
FT CARBOHYD 99 99  
FT CARBOHYD 159 159  
FT CARBOHYD 271 271  
FT CARBOHYD 319 319  
FT SEQUENCE 397 AA; 45186 MW; 97A38595A0523947 CRC64;  
Query Match 22.2%; Score 486; DB 1; Length 397;  
Best Local Similarity 33.4%; Pred. No. 1.9e-34;  
Matches 124; Conservative 65; Mismatches 154; Indels 28; Gaps 14;  
QY 49 QLLPLPGYPTCHNVTGKDFLLSLQHPHG-KNKAADSTGPPVFLQ--HGLFQGGDTWF 105  
DB 37 EIIMHWGYP--EHSVQTGDIYLVGHVRPHGRKNQFDKGPVYVLOWRHGFLADSSNV 94  
QY 106 INSAEOSLYIADNAGFDYWIGNVTRKSGHSTFVHDKLFWDWSWELABYDLLAML 165  
DB 95 TFINNNLSFLIADAGDFVMGNSRGTSKHKTLVSQDEYWFASFDEMAKYDLPASI 154  
QY 166 GVVYTVTSQKILY-VGHSOGTGMGLAALN-MPEIVKMISSAALLCPISYLDHVSASFVLR 223  
DB 155 NVILNKTGQELNVGHSOGCTIGFTAFSQMPELAKVMFFALAPVLSNFASGPMVKL 214  
QY 224 AVAMHLDQMLV--TMGIHQLNFRSDMGVQIVDSLCDGEHVD-----CNLLLSAITGEN-CC 276  
DB 215 G---RLPDLLEDLFQKQFLQPSAMVKKLSTHIC--THVMKELCANIFFLICGFNEKN 269  
QY 277 FNTSRDLYLETPHPSSTKNLHLFQMTKGTFAKYDYGLLG-NLRRYGHURLPPAFDLS 335  
DB 270 LNNRSVDVYTHCPAGTSVQNNVHWTVQVYKYLQAFDWSGSSDKNFYHYNQSYPLYSIK 329  
QY 336 SIPESLPIWNGYGLDALADVTDVORTIRELGSTPELLY---IGDYGHIDFVMSVAKDD 392  
DB 330 DM--QLPTALWGGKDWLADTSDINILLTEI---PPLVYHKNIPEMDHLDIFWGLDAPWR 384

QY 393 VYVDLIRFLRE 403  
DB 385 LYNEVVSMLMK 395  
RESULT 9  
LIP1\_DROME  
ID LIP1\_DROME STANDARD; PRT; 439 AA.  
AC O46107: Q9VKR6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Lipase 1 precursor (EC 3.1.1.1-).  
GN LIP1 OR CG7279.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN (1)  
RP SEQUENCE OF 7-433 FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=98227315; PubMed=9566193;  
RA Pistillo D., Manzi A., Tino A., Pilo Boyl P., Graziani F., Malva C.;  
RT "The Drosophila melanogaster lipase homologs: a gene family with  
RT tissue and developmental specific expression.";  
RL J. Mol. Biol. 276:877-885(1998).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abil J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., CWiley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Dey I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E.C., Stadling A.C., Stapleton M., Strong R., Sun E.,  
RA Swigras R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: COULD BE A DIGESTIVE ENZYME.  
CC -!- TISSUE SPECIFICITY: OVARIES.  
CC -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.). ALSO SIMILAR TO

```
CC -----
CC LEPIDOPTERAN EGG-SPECIFIC AND YOLK PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y14366; CAA74736.1; -.
DR EMBL; AE003629; AAF52994.1; -.
DR FlyBase; FBgn0033496; Lip1.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00561; abhydrolase; 1.
DR PROSITE; PS00120; LIPASE_SER; FALSE_NEG.
KW Hydrolase; Lipid degradation; Signal; Glycoprotein.
FT SIGNAL 1 24
FT CHAIN 25 439
FT DOMAIN 30 44
FT ACT_SITE 197 197
FT ACT_SITE 393 393
FT CARBOHYD 124 124
FT CARBOHYD 151 151
FT CARBOHYD 346 346
FT CARBOHYD 379 379
FT CARBOHYD 426 426
FT CONFLICT 10 10
FT CONFLICT 213 213
FT CONFLICT 412 412
FT SEQUENCE 439 AA; 50660 MW; 9E32E20BEAE93E3F CRC64;

Query Match 19.58; Score 427; DB 1; Length 439;
Best Local Similarity 30.18; Pred. No. 2.7e-157;
Matches 110; Conservative 67; Mismatches 157; Indels 32; Gaps 12;

Qy 48 DQLLLPLGPGCTEHNVTETKDGFLSLQHPHGKKAADSTGPPVFLQHLFGQGGDTWFIN 107
Db 69 DKLIAGYGESEVHHVTTEDGYLTHMRI---RKQGA---PPFLQHLGVDSSAGFVVM 121
Qy 108 SAEOSLGYLADNGFDVWGNVGRTRWSKGHSTFSVHDKLFDWWSQBLAEYDILLMLGY 167
Db 122 GPNVSLAYLLADHNDVWLGNAAGNRYSRNHTLDPDESKFWSHWEIGMYDLPAMTDH 181
Qy 168 YTVTQ-SKILYVGHSGQTGIMGLAALTW-PEIVKMISAAALCPISYLDHVSASFVLRV 225
Db 182 VLKVTGFPKPLHYAGHSQSGCTFFVCMRPAAYNDKVVSMQALAPAYAKETEDHPYIRAI 241
Qy 226 AMHLDQMLVTGIIHOL---NFRSDMGVQIVDSLC---DGEHVDCNNLLSAITGENCC-F 277
Db 242 SLVFNLS-LVGSSIREMFNGEPR-----FLCRMTETEEL-CLEAVFGIVGRNWNNEF 290
Qy 278 NPSRIDYLEYEPHPSSTKNLHLLFQMIKGTAKFYGLLGNRLRYGHLRPPAPDLSSI 337
Db 291 NRKMFPVTLGHYPAGAAQVKHFTIILKSGRFAPYSYSSNKNMQLYRDLHLPYRNLSLV 350
Qy 338 PSLPIWMGYGGDLADADVTDVQRTIRELST--PELLYIGDYGHIDFVMSKADDDVYV 395
Db 351 --TVPTFVYTSNDLLCHPKDVESMCDLGNVTGKYLVPQKEFNMHDLWADIVRKMLYR 408
Qy 396 DLIRFL 401
Db 409 RMLQVL 414

RESULT 10
TGILL YEAST
ID TGILL YEAST STANDARD; PRT; 548 AA.
AC P34163.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Triglyceride lipase-cholesterol esterase (EC 3.1.1.-).
GN TGL1 OR YKL140W OR YKL5.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92245761; PubMed=1574929;
RA Abraham P.R., Mulder A., Van't Riet J., Planta R.J., Raue H.A.;
RT "Molecular cloning and physical analysis of an 8.2 kb segment of
RT chromosome XI of Saccharomyces cerevisiae reveals five tightly linked
RT genes.";
RL Yeast 8:227-238(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX Red M.R., Xu G., Kirchraeth L., Fritz C., Keuchel H., Hollenberg C.P.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z25464; CAA80958.1; -.
DR EMBL; Z28140; CAA81981.1; -.
DR PIR; S37969; S37969.
DR PIR; S39000; S39000.
DR SGD; S0001623; TGL1.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00561; abhydrolase; 1.
KW Hydrolase; Lipid degradation.
SQ SEQUENCE 548 AA; 62979 MW; 32D1F230701CB083 CRC64;

Query Match 13.48; Score 293.5; DB 1; Length 548;
Best Local Similarity 24.88; Pred. No. 1.2e-17;
Matches 95; Conservative 67; Mismatches 136; Indels 85; Gaps 16;

Qy 60 EHNVTETKDGFLSLQHPHGKKAADSTGPPVFLQHLFGQGGDTWFIN-SAEOSLGYL 117
Db 79 DHLVRTEDNYILTTHRIPPISKRNKKV---VYLHGLLSCSDVWCNIEHKNLPVL 135
Qy 118 ADNGFDVWGNVGRTRWSKGHSTFSVHDKLFDWWSQBLAEYDILLMLGYVTVTQ-SKI 176
Db 136 HDLGYDVMGNRGNKYSTAHLNPKPKSNKEWDFSIDFAFDIPNSIEFDITDKVKV 195
Qy 177 LYVGHSGQTGIMGLAALTWPEIVKMISAAALCPISYLDHVSASFVLRVAVMHLDQMLVTM 236
Db 196 ICIGFSQSAQMAFAFSLSE-----KLARKVSHFIAIAPAM-----TPK 234
Qy 237 GIHQLNFRSDMGVQIVDSLCDE-----HVDCCNNLLSAITG 272
Db 235 GLHN-----RIVDTLAKSGPGMYLFFGRKIVLPSPAVIQWTLHPTLFLNCIDIAN 285
Qy 273 ENCCPN-----TSRIDYLEYEPHPSSTKNLHLLFQMIKGTAKFYDY--GLLGNLR 322
Db 286 K-ILFNWKSFNILPQKTIASYAKLY-STTSVKSVIWHFQILRSQKQFOMFEESDNMLSLT 343
Qy 323 RYGHLPAPPAFDLSSTPE----SLPIWMGYGGDLADADVTDVQRTIRELGSYPELLYIGDY 378
Db 344 R-----PYQIANPPTTNIKIPILLIYGGIDSLVDI-DVWKNLPPNSVFD-VKVDNY 394
Qy 379 GHIDEVMSVKAKDDVYDLIRFL 401
Db 395 EHLDLGWKADTLVIKVLRFI 417
```



```
RESULT 11
YAGX_ECOLI
ID YAGX_ECOLI STANDARD; PRT; 841 AA.
AC P77802;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yagX precursor.
GN YAGX OR B0291.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=562;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [12]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Korman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO E.COLI PLASMID NTP513 CFA FIMBRIA SUBUNIT
CC CFA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000136; AAC73394.1;
CC EMBL; U73857; AAB18020.1;
CC ECoGene; EG13563; yagX.
CC KW Hypothetical protein; Signal; Complete proteome.
CC FT SIGNAL 1 29 POTENTIAL.
CC CHAIN 30 841 HYPOTHETICAL PROTEIN YAGX.
CC SEQUENCE 841 AA; 91228 MW; D2016BB0ACD726AC CRC64;
Query Match 4.9%; Score 107.5; DB 1; Length 841;
Best Local Similarity 22.3%; Pred. No. 0.24;
Matches 81; Conservative 47; Mismatches 132; Indels 103; Gaps 18;
QY 65 TKDGLLSLQHPCHKNAADSTGPPVFLQHLGFGGDTWFIN-----SA 109
DB 300 TRDGRLLSVQNTMG-NHEVDTRGLP-----YGIYDVEVIVNGVRISKRTQVKNLFSR 354
QY 110 EOSLGYILADNGDFVWIGNVGRTRSK-----GHSTFSVHDKLFWDMWSOE 155
DB 355 GRGVGAPLA---WQWGGSFHMDRWSKCKTRPAKESWLAGASTSGSLTSLW----- 405
QY 156 LAEYDLLAMGLGVYVYVTSKILYVGHSGQTIMGALATMP-----EIVKMISAAALC 208
DB 406 -----AATGYCY-----DNQAVGETRLTLPGLGAINVNLQNLAADS----- 442
QY 209 PTISYLDHVSASFVRAVAMHLDQMLVTMGHOLNFRSDMGVQIVDSLCDGHEHVDNCNLLS 268
DB 443 SWSISGISATNLPGGFSSLUWVQKTRIG-NQLR-RSD-----ADNRAIGGTNLNLSLWS 495
QY 269 AITGNCNCTSRIDYILEYEPHPSSTKNLHFLQMKRGTFKADYGLGNLRY--GH 326
DB 496 KLGTSISYNDNR-----RYNSHYTA---DYQNVSSTFG--SLGLRAGIQRYNNDG 544
QY 327 LRPPAFDLSSIPESLP--WM-----GYGGLDALADVTQVQRTIRELGSTPELLYIG 376
```

```
DB 545 SNANTGKYIALDLSPLGNWFSAGMTHQNGYTMANLSARKQFDEGTIRTVGANLSRAISG 604
QY 377 DYX 379
DB 605 DTG 607
RESULT 12
YFX5_CAEEL
ID YFX5_CAEEL STANDARD; PRT; 1216 AA.
AC Q09277;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 133.3 kDa protein F40H6.5 in Chromosome III.
GN F40H6.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: STRONG, TO C.ELEGANS F40H6.2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U21310; AAA62522.1;
CC WormPep; F40H6.5; CE01937.
CC InterPro; IPR004024; DUF225.
CC DR InterPro; IPR001079; Galectin.
CC DR Pfam; PF02795; DUF225; 1.
CC SMART; SM00276; GLECT; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 3 23 POTENTIAL.
CC TRANSMEM 181 201 POTENTIAL.
CC TRANSMEM 377 397 POTENTIAL.
CC TRANSMEM 440 460 POTENTIAL.
CC TRANSMEM 781 801 POTENTIAL.
CC TRANSMEM 1174 1194 POTENTIAL.
CC SEQUENCE 1216 AA; 133313 MW; B5C30EBDA014D793 CRC64;
Query Match 4.6%; Score 100; DB 1; Length 1216;
Best Local Similarity 22.8%; Pred. No. 1.7;
Matches 52; Conservative 33; Mismatches 89; Indels 54; Gaps 9;
QY 161 LLAMLGYVTVYVTSKILYVGHSGQTIMGALATMPDIVKMISSAALLCPSYLDHVSASF 220
DB 6 LLFLVLSTLVESVYQY-----QILLEPIRYFTDSAAVPAKSIYFTNLPQH 54
QY 221 VL--RAVAMHLDQMLVTMGHOLNFRSDMGVQIVDSLCDGHEHVDNCNLLSAITGENCCFN 278
DB 55 SLTKSSKLDLSQFNVRVGRFTQEMFLKNLG-----CGTD--TCQNFATVTS---VYN 102
QY 279 TSRIDYILEYEPHPSSTKNL-----HHLFQMIKRKCTFAK-----YDGLGN-- 320
DB 103 IIDLGRKLFETNPACTDNVLGKAGYLAETEGLRVDEYSEEDDSTFYNGKYPNQF 162
QY 321 ----LRRYGHLR-----PPAFDLSSIPESLPIMWGYGGLDALADVT 357
DB 163 GACQLRDYGLKRYKFVRTLGHAFATVAIPENLPISYIFAYDGLPEPT 210
```

RESULT 13  
ID UREL\_SYN3 STANDARD; PRT; 569 AA.  
AC P73061;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase).  
GN UREC OR SLL1750.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).  
CC -!- COFACTOR: Binds 2 nickel ions per subunit (By similarity).  
CC -!- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).  
CC -!- PTM: Lys-219 is carbamylated. The carbamoyl group provides the  
CC ligands for the two nickel ions (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE UREASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90903; BAA17083.1; -.  
DR HSSP; P18314; 1FWE.  
DR MEROPS; M38.0NW; -.  
DR InterPro; IPR001924; UreaseA.  
DR Pfam; PF00449; urease; 1.  
DR Pfam; PF02802; urease; 1.  
DR ProDom; PD002467; UreaseA; 1.  
DR PROSITE; PS00145; UREASE\_2; 1.  
DR PROSITE; PS01120; UREASE\_1; 1.  
KW Hydrolase; Metal-binding; Nickel; Complete proteome.  
FT METAL 136 136 NICKEL 2 (BY SIMILARITY).  
FT METAL 138 138 NICKEL 2 (BY SIMILARITY).  
FT METAL 219 219 NICKEL 1 AND 2 (BY SIMILARITY).  
FT METAL 248 248 NICKEL 1 (BY SIMILARITY).  
FT METAL 274 274 NICKEL 1 (BY SIMILARITY).  
FT METAL 362 362 NICKEL 2 (BY SIMILARITY).  
FT ACT\_SITE 322 322 BY SIMILARITY.  
SQ SEQUENCE 569 AA; 61037 MW; 1C720A254335D7FD CRC64;  
  
Query Match 4.4%; Score 97.5; DB 1; Length 569;  
Best Local Similarity 19.6%; Pred. No. 1.1; 135; Indels 105; Gaps 16;  
Matches 71; Conservative 51; Mismatches 51;  
  
QY 77 PHGKNKAADSTGPPVFLQHLFGQGGDTWFINSAEQSLGYLADNGFDWIGNVGRTRWSK 136  
DB 163 PRACTKATCTT-PGAWNTHRMLOADGFPIN-----LGLFGKNG-----SQPAALAE 209  
QY 137 GHSTFSVHDKLFWDMSQELA-----EYDL-----LAMLGYVYT---VTQSKI 176  
DB 210 QIKAGAIGLKLHEDWCTTPAADIICLGVADYDVQVAIHTDTLNEAGFVETTTAALKNRY 269  
QY 177 LVVGHSGQITMGLAALTMTPEIVKMISSALLCPISYLDHVSASFVLRAVAMHLDQMLVTM 236  
DB 270 IHTYHTGAGGGHA-----PDIKICGELNVL-PSS--TNPTRPYTINTLEEHLDMLVCH 322

QY 237 GIHQLNFRSDMGVOIVDLSLDCGHEHVCNNLLSAI----- 270  
DB 323 HLHR-NIPED--VAFAESRIRRETIAEDILHDLGAFSIISSDQAMGRVGEVICRTWOT 379  
QY 271 -----TCENCCFNSTRIDYLEYEPHPSSTKNLHFLQMIKRGTF----- 311  
DB 380 AHKMKVQRGOLPGETGNNDNLRAKR--YVAKYITINPAITHGISEEIGSVVEGKLADLCIW 437  
QY 312 -----KYDYGLLGNLRRYGHLPAPAFDLSLSSIPESLPIWN-----GYGGLDALADYTDV 359  
DB 438 SPAFFGVKPELVIKGGIVAVQAMCDAN--ASITPPQVHMQPMFANYGGAIAATSVTFF 494  
QY 360 QR 361  
DB 495 SQ 496  
  
RESULT 14  
LIPP\_HORSE  
ID LIPP\_HORSE STANDARD; PRT; 461 AA.  
AC P29183;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Triacylglycerol lipase, pancreatic precursor (EC 3.1.1.3) (Pancreatic  
DE lipase) (PL) (Fragment).  
GN PNLIIP.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=92267022; PubMed=1587279;  
RA Kerfelec B., Foglizzo E., Bonicel J., Bougis P.E., Chapus C.;  
RT "Sequence of horse pancreatic lipase as determined by protein and  
RT cDNA sequencing. Implications for p-nitrophenyl acetate hydrolysis by  
RT pancreatic lipases.";  
RL Eur. J. Biochem. 206:279-287(1992).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=94238688; PubMed=8182745;  
RA Bourne Y., Martinez C., Kerfelec B., Lombardo D., Chapus C.,  
RA Cambillau C.;  
RT "Horse pancreatic lipase. The crystal structure refined at 2.3-A  
RT resolution.";  
RL J. Mol. Biol. 238:709-732(1994).  
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X66218; CAM46961.1; -.  
DR PIR; S21251; S21251.  
DR PIR; S21223; S21223.  
DR PDB; 1HPL; 31-MAY-94.  
DR InterPro; IPR000734; Lipase.  
DR InterPro; IPR001024; Lipoxigenase\_LH2.  
DR InterPro; IPR000379; Ser\_estrif\_site.  
DR Pfam; PF00151; lipase; 1.  
DR Pfam; PF01477; PLAT; 1.  
DR SMART; SM00308; LH2; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
KW Hydrolase; Lipid degradation; Pancreas; Signal; 3D-structure.



Search completed: December 30, 2002, 20:44:30  
Job time : 41 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2002, 20:09:51 ; Search time 92 Seconds  
(without alignments)  
918.253 Million cell updates/sec

Title: US-09-699-652A-14  
Perfect score: 2194  
Sequence: 1 MAMAGHAPGALPLILLWS.....DDVYVDFLRENGWHNSY 410

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhnc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1055	48.1	344	10 Q9SKL5	Q9skl5 arabidopsis
2	571.5	26.0	684	5 Q95XV1	Q95xv1 caenorhabdi
3	557.5	25.4	403	5 Q61866	Q61866 caenorhabdi
4	549.5	25.0	404	5 Q16936	Q16936 caenorhabdi
5	549	25.0	405	5 Q93789	Q93789 caenorhabdi
6	540	24.6	318	10 Q9FMT3	Q9fmt3 arabidopsis
7	539	24.6	395	11 Q9D798	Q9d798 mus musculu
8	539	24.6	395	11 Q9D611	Q9d611 mus musculu
9	538	24.5	395	11 Q9D7C5	Q9d7c5 mus musculu
10	538	24.5	395	11 Q9D6P3	Q9d6p3 mus musculu
11	537	24.5	395	11 Q9D767	Q9d767 mus musculu
12	537	24.5	395	11 Q9D6X0	Q9d6x0 mus musculu
13	537	24.5	395	11 Q9D6T5	Q9d6t5 mus musculu
14	537	24.5	395	11 Q9D6S5	Q9d6s5 mus musculu
15	537	24.5	395	11 Q9D6N8	Q9d6n8 mus musculu
16	537	24.5	395	11 Q9C9P7	Q9c9p7 mus musculu

17	536	24.4	395	11 Q9D6O6	Q9d6o6 mus musculu
18	536	24.4	395	11 Q9D6Q2	Q9d6q2 mus musculu
19	535	24.4	395	11 Q9D796	Q9d796 mus musculu
20	535	24.4	395	11 Q9D6L9	Q9d6l9 mus musculu
21	531	24.2	395	11 Q9D6Q3	Q9d6q3 mus musculu
22	530	24.2	395	11 Q9D766	Q9d766 mus musculu
23	530	24.2	395	11 Q9C9P8	Q9c9p8 mus musculu
24	529	24.1	395	11 Q9D6M9	Q9d6m9 mus musculu
25	528	24.1	395	11 Q9D760	Q9d760 mus musculu
26	502	22.9	374	4 Q96LG2	Q96lg2 homo sapien
27	500.5	22.8	426	5 Q17766	Q17766 caenorhabdi
28	498	22.7	456	5 Q9VKS5	Q9vks5 drosophila
29	495.5	22.6	434	5 Q9VKY9	Q9vkt9 drosophila
30	474	21.6	411	5 Q20449	Q20449 caenorhabdi
31	470.5	21.4	351	5 Q95X33	Q95x33 caenorhabdi
32	468.5	21.4	399	5 Q9VPE9	Q9vpe9 drosophila
33	466.5	21.3	411	5 Q94252	Q94252 caenorhabdi
34	458	20.9	398	5 Q9V796	Q9v796 drosophila
35	425.5	19.4	838	5 Q9VBR6	Q9vbk6 drosophila
36	418	19.1	416	5 Q9VQ05	Q9vqq5 drosophila
37	417	19.0	457	5 Q9VKT2	Q9vkt2 drosophila
38	417	19.0	457	5 Q8T3X7	Q8t3x7 drosophila
39	416.5	19.0	559	5 Q17219	Q17219 bombyx mori
40	412.5	18.8	406	5 Q9VKT7	Q9vkt7 drosophila
41	410.5	18.7	616	5 Q77107	Q77107 plodia inte
42	401	18.3	504	5 Q94568	Q94568 galleria me
43	392.5	17.9	401	5 Q9U276	Q9u276 caenorhabdi
44	384.5	17.5	355	5 Q9VKT0	Q9vkt0 drosophila
45	379	17.3	435	5 Q9VG46	Q9vg46 drosophila

## ALIGNMENTS

RESULT 1

ID	Q9SKL5	PRELIMINARY:	PRT:	344 AA.
AC	Q9SKL5;			
DT	01-MAY-2000 (TRENBLrel. 13, Created)			
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)			
DE	01-MAR-2002 (TRENBLrel. 20, Last annotation update)			
DE	Putative lysosomal acid lipase.			
GN	AT2G15230.			
OS	Arabidopsis thaliana (Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RC	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,			
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,			
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.,			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RT	thaliana.";			
RL	Nature 402:761-768(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Lin X.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC006238; AAD25569.1;			
DR	InterPro; IPR000734; Lipase.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	PROSITE; PS0120; LIPASE_SER; 1.			
DR	SEQUENCE 344 AA; 38868 MW; 7BC4ED067E3602C4 CRC64;			

Query Match	48.1%	Score 1055;	DB 10;	Length 344;
Best Local Similarity	54.5%	Pred. No. 1.2e-88;		
Matches 196;	Conservative 42;	Mismatches 71;	Indels 52;	Gaps 2;
QY 40	GS GSGGLCDQLLLPLGYPCTEHNHVTETKDGFTLLSLQHIPHGKKNKAADSTGPPVFLQHGLFQ	99		
Db 26	GSPVNSLCADLIHPANYSCTEHSIQTQDGYTLAQRVASLGRLL--OSGPPVLLQHGLFM	83		
QY 100	GGDTWFINSAQSGLGYTLADNGDFDWIGNVRGTRWSKGHSTFSVHDKLFDWDSWOELAEY	159		
Db 84	AGDWFLNSPRESLGFTLADHGFVWVGNVGRGTRYSGHVLTSDTDKRFWDWSQDLAMY	143		
QY 160	DL LAMLGXYVTVTQSKILYVGHSGCTIMGLAALTMPEIVKMMISSAALCPISYLDHVSAS	219		
Db 144	DLAEMIQLYISISKSKITFLVGHSGCTIMSPAALTQPHVAEMVEAAALCPISYLDHVTAP	203		
QY 220	FVLRVAVHMLDQMLVTWGIHQNLFRSDMGVQIVDSLCDGEHVDNLLSALTGCNCCENT	279		
Db 204	LVERVFMHLQD-----	215		
QY 280	SRIDYLEYEPHPSSTKNLHHLFQMIKRGTEAKYDYGLLGNLRRYGHRLRPPAFDLSSIP	339		
Db 216	--IEYLDYEPHPSVKNIHFLQMIKRGTEAQDYGYFKNLRYGLSKPPEFTILSHIP	273		
QY 340	SLPIWMVGGLDALADYTDVQRTIRELGSTPELILYIGDYHDFVMSYKAKDDVYVDLIR	399		
Db 274	SLPMMVGYGTDGLADYTDVETLAELPSPPELLYLEDYGHIDFVLGSSAKEDVYKHM IQ	333		
QY 400	FLR 402			
Db 334	FFR 336			

RESULT 2

Q95XV1

ID

Q95XV1

PRELIMINARY;

PRT;

684 AA.

AC

Q95XV1;

DT

01-DEC-2001 (TremBLrel. 19, Created)

DT

01-DEC-2001 (TremBLrel. 19, Last sequence update)

DE

01-MAR-2002 (TremBLrel. 20, Last annotation update)

DE

Hypothetical 77.3 kDa protein.

GN

Y57E12B.3.

OS

Caenorhabditis elegans.

OC

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC

Rhabditidae; Peloderinae; Caenorhabditis.

OX

NCBI\_TaxID=6239;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=BRISTOL N2;

RX

MEDLINE=99069613; PubMed=9851916;

RA

None;

RP

SEQUENCE FROM N.A.

RC

STRAIN=BRISTOL N2;

RA

Becker M.;

RT

"The sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";

RL

Science 282:2012-2018(1998).

[2]

RN

RN

RP

SEQUENCE FROM N.A.

RC

STRAIN=BRISTOL N2;

RA

Becker M.;

RT

"The sequence of C. elegans cosmid Y57E12B.";

RL

Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.

[3]

RN

RN

RP

SEQUENCE FROM N.A.

RC

STRAIN=BRISTOL N2;

RA

Waterston R.;

RT

"Direct submission.";

RL

Submitted (NOV-2001) to the EMBL/GenBank/DBDJ databases.

DR

EMBL; AC024835; AAK68537.1; -

DR

InterPro; IPR00073; Abhydrolase.

DR

InterPro; IPR000734; Lipase.

DR

InterPro; IPR000379; Ser\_estrs\_site.

DR

Pfam; PF00561; abhydrolase; 1.

DR

PROSITE; PS00120; LIPASE\_SER; UNKNOWN\_1.

KW

Hypothetical protein.

QY 49 QLLLPLGYPCTEHNVETKDGFLLSLQHIPHGKNAADSTG--PPVFLQHCLFQGGDTWFI 106

```
Db 33 QIIMRWGYPAMIYDVTEDGYILELHRIPIYKTNWTPNGKKPVVFMQHGLECSSSNVV 92
Qy 107 NSAQSLGYILLADNGFDVWICNVGRTRWSKGHSFVSHDKLFWDSWQELAEYDLAMLG 166
Db 93 NLPTESAAFLADAGYDVLGNFGNTYSMKHKNLKPSSHSAFWSWDEMOQYDLPAMIE 152
Qy 167 YVYTVT-OSKILYVGHSGQTMGLAALTMPET--VKMISSAALLCPISYLDHVSAS--FV 221
Db 153 KALEVTGDSLYYIGHSGQTLTMFSRLSEDKVGNKIKKFFALAPVGSVKHIKALKFF 212
Qy 222 LRAVAMHLDQMLVTMGIHQLNFRSDMGVQIVDSLCDGHEVD---CNNLSAITG-ENCCF 277
Db 213 ADYFSLEDFGDFVDFGSEFLFNWIMKLVSFCAGLKVEAGVCDVDMFELIAGPESQL 272
Qy 278 NTSRTDYLYEPHPSSPKNLHLHFLQMTKCTFAKYDYGGLGNLRRYGHRLRPPAFDLSSI 337
Db 273 NATRPVIVAHTPAGTSTQNTVHVIQWVRHGGTPKYDGEKNKHKYGOANVPAYDFITV 332
Qy 338 PESLPIMWGYGGLDALADVTQVORTIRELSTPELL-----YIGDYGHIHDFVM 385
Db 333 --NRPVLYWGDSDLADPTDV-----TDFLLTHLNFSTVVQNNKLIDYNHLDFIW 381
Qy 386 SVKAKDDVYVDLIRFLRN 404
Db 382 GLRAPKDIYEPIIDVRND 400

RESULT 6
Q9FMT3 PRELIMINARY; PRT; 318 AA.
AC Q9FMT3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE Similarity to lipase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones."
RL DNA Res. 4:401-414(1997).
DR EMBL; AB007650; BAB08297.1;
SQ SEQUENCE 318 AA; 36448 MW; 25D72693E454743F CRC64;

Query Match 24.6%; Score 540; DB 10; Length 318;
Best Local Similarity 35.3%; Pred. No. 2.7e-41;
Matches 118; Conservative 54; Mismatches 108; Indels 54; Gaps 8;

Qy *103 TWFINSAQSLGYILLADNGFDVWICNVGRTRWSKGHSFVSHDKLFWDSWQELAEYDLL 162
Db 2 SWLLPADQNLPLLIADNGFDVWICNVGRTRWSKGHSFVSHDKLFWDSWQELAEYDLLP 61
Qy 163 AMLVGYVTVTSQKILYVGHSGQTMGLAALTMPETV---KMISSAALL-----CPIS 211
Db 62 AMFDHIHGLTGKIHLYLGHs-----LVGPTICFVLRKRVGSSIGGDVSRCLSQ 112
Qy 212 YLDH-----VSASFVLRVAMHLDQMLVTMGIHQLNFRSDMGVQIVD 253
Db 113 PHDHRHRRYRKNLPRRVKMKITCLSSSEKNRP-PLFDQGLSTRKVG-----DRK 161
Qy 254 SICDGEHVDNCNLLSAITGNCCTNTSRIDYLYEYEPHPSKTNLHLFQMRKGTFAKY 313
Db 162 AICLKAGIDCVLDSVITGKNCCNLNASTIDFLANEPSTSTKNIHIAQTVRKELRY 221
```

```
Qy 314 DYGLLG-NLRRYGHRLRPPAFDLSSIPESLPIMWGYGGLDALADVTQVORTIRELG----S 368
Db 222 NYGSSDRNKRHYGOAIPAYNISAIPELPLFFSYGGLDSLADVKDFEFLDQKRYHDID 281
Qy 369 TPPELLYIGDYGHIDFVMSVAKDDVYVDLIRFLR 402
Db 282 KMNQVQKDYAHADFIMGVTAQDVVYNQVATFEK 315

RESULT 7
Q9D798 PRELIMINARY; PRT; 395 AA.
AC Q9D798;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE 2310051B2IRIK protein.
GN 2310051B2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynchaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK009431; BAB26283.1;
DR MGD; MGI:1914967; 2310051B2IRIK.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00561; abhydrolase_1.
DR PROSITE; PS00120; LIPASE_SER; 1.
SQ SEQUENCE 395 AA; 44603 MW; D3FD8B86FEA671E3E CRC64;

Query Match 24.6%; Score 539; DB 11; Length 395;
Best Local Similarity 33.6%; Pred. No. 4.6e-41;
Matches 129; Conservative 68; Mismatches 135; Indels 52; Gaps 13;

Qy 49 QLLPLGYPCTEHNVTYKDGELLSLQIHPIHGKNAADSTG--PPVFLOHGLFQGGDTWFI 106
Db 36 QMITWGWPSYEYEVTEEDGYILGVYRIPYK-KNSENIGKRPVAYLQHLQIASATNWT 94
Qy 107 NSAQSLGYILLADNGFDVWICNVGRTRWSKGHSFVSHDKLFWDSWQELAEYDLAMLG 166
Db 95 NLPNNSLAFLIADAGYDVLGNFGNTYSMKHKNLKPSSHSAFWSWDEMOQYDLPAMIE 154
Qy 167 YVYTVT-OSKILYVGHSGQTMGLAALTMPET--VKMISSAALLCPISYLDHVSAS----- 218
Db 155 FIVOKTGOEKTHYVGHSGQTTIGTIAFTSNPALAKKIRFALAPVATVKTYESPFKIS 214
Qy 219 ---SEVLRAV-----AMHLDQMLVTMGIHQLNFRSDMGVQIVDSLCDGHEVD---CNN 265
Db 219 ---SEVLRAV-----AMHLDQMLVTMGIHQLNFRSDMGVQIVDSLCDGHEVD---CNN 265
```



```

Db 215 LIPKFLKLVIFGNKMPHNYLDQFLGT-----EVCSRELLDLCSN 256
QY 266 LLSAITG-ENCCFNTSRIDYILEYEPHPSSTKNLHFLFOMIRKGTFAKYDYL-LGNLRR 323
Db 257 ALFICGFDKKNLNSRLDVLGHNPAGTSTQDLFWHAQLAKSGKLOAYNWGSPQLNMLH 316
QY 324 YGHLRPPAFDLSSIPESLPIMWGYGGLDALADVTQRTIRELGSTPELly----IGDYGH 380
Db 317 YNOKTPPYDVSAMTVPIAVW--NGGHDILADPQDVAMLLPKL---PNLLYHKELLYPNH 371
QY 381 IDFMVSVKAKDDVYVDLIRFLREN 404
Db 372 LDFIAMDAPQEVYNEIVTMAED 395

RESULT 8
Q9D6L1
ID Q9D6L1 PRELIMINARY; PRT; 395 AA.
AC Q9D6L1
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Adult male tongue cDNA, RIKEN full-length enriched library,
clone:2310079020, full insert sequence.
GN 2310051B21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaiji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK010231; BAB26784.1;
DR MGD: MGI:1914567; 2310051B21RIK.
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000734; Lipase.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00561; abhydrolase; 1.
DR PROSITE: PS00120; LIPASE_SER; 1.
SQ SEQUENCE 395 AA; 44659 MW; 130BDB95FC5A40E0 CRC64;

Query Match 24.6%; Score 539; DB 11; Length 395;
Best Local Similarity 33.9%; Pred. No. 4.6e-41;
Matches 130; Conservative 66; Mismatches 136; Indels 52; Gaps 13;

QY 49 QLLPLPGYCTEHNVTGDKGFLLSLQHIPGKNKAADSTG--PPVFLQHLFGGDTWFI 106
Db 36 QMITWYGPSEYEVVETEDGYILGVYRIPYK-KNSENIGKRPVAYLQHLIASATNWT 94
QY 107 NSAEQSLGVLADNGFDWIGNVGRTRWSKGHSFVHDKLFWDWSWQBLAEYDLLAMLG 166
Db 95 NLPNNSLAFILADAGYDVLGNRGNTWSRKNVYSPDSVEFWAFSDEMAYIDLPAITD 154

```

```

QY 167 YVYTVT-QSKILYVGHSGQTIMGLAAL-TWPEIVKVMISSAALICPISYLDHVS----- 218
Db 155 FIVQKTGQEKIHYVGHSGQTTIGTAFSTNPALAKKIKRFYALAPVATYKTESPFKKIS 214
QY 219 ---SFVLRV-----AMHLDQMLVTWGIHQINFRSDMGVOIVSLDCGHDV---CNN 265
Db 215 LHKFLKLVIFGNKMPHNYLDQFLGT-----EVCSRELLDLCSN 256
QY 266 LLSAITG-ENCCFNTSRIDYILEYEPHPSSTKNLHFLFOMIRKGTFAKYDYL-LGNLRR 323
Db 257 ALFICGFDKKNLNSRLDVLGHNPAGTSTQDLFWHAQLAKSGKLOAYNWGSPQLNMLH 316
QY 324 YGHLRPPAFDLSSIPESLPIMWGYGGLDALADVTQRTIRELGSTPELly----IGDYGH 380
Db 317 YNOKTPPYDVSAMTVPIAVW--NGGHDILADPQDVAMLLPKL---PNLLYHKELLYPNH 371
QY 381 IDFMVSVKAKDDVYVDLIRFLREN 404
Db 372 LDFIAMDAPQEVYNEIVTMAED 395

RESULT 9
Q9D7C5
ID Q9D7C5 PRELIMINARY; PRT; 395 AA.
AC Q9D7C5
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE 2310051B21RIK protein.
GN 2310051B21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaiji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK009359; BAB26240.1;
DR MGD: MGI:1914567; 2310051B21RIK.
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000734; Lipase.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00561; abhydrolase; 1.
DR PROSITE: PS00120; LIPASE_SER; 1.
SQ SEQUENCE 395 AA; 44748 MW; 3ED43992458DE058 CRC64;

Query Match 24.5%; Score 538; DB 11; Length 395;
Best Local Similarity 34.1%; Pred. No. 5.7e-41;
Matches 131; Conservative 65; Mismatches 136; Indels 52; Gaps 14;

QY 49 QLLPLPGYCTEHNVTGDKGFLLSLQHIPGKNKAADSTG--PPVFLQHLFGGDTWFI 106

```

Db	36	QMTITYGYPSEEEYVYVTDGYTLGVYRIPYGK-KNSENIGKRPVAYLQHLGIASAKNMIT	94
Qy	107	NSAEOSLGYTLADNGFDVWIGNVGRTRWSKGHSTGTSVHDKLFWDMSWOELAEYDILLMLG	166
Db	95	NLPNNSLAFIADAGYDVWLGNSRGNTWSKKNVYSPDSVEFAWFSFDEMAKYDLPAID	154
Qy	167	YVYTVT-QSKILYVGHSGQGTIMGLAAL-TMPEIVKMISSAALLCPIS---YLD-----H	215
Db	155	FIVQKTQGEKIHVGHSGQGTIGTIFAFSTNPALAKKIKRFYALAPVATVKYTESPFKKIH	214
Qy	216	VSASFVLRAV-----AMHLDQMLVTWGIHQLNFRSDMGVQIVDSLCDGSHVD-CNN	265
Db	215	FIPKFLKVFIFGNKMFPHNYLDQFLGT-----EVCSERLDLLCSN	256
Qy	266	LLSAITG-ENCCENTSRIDYILEVEPHPSSTKNLHLFOMIRKGTFAKYDYG-LGNLRR	323
Db	257	ALFIFCGFDKKNLVNRFEDVYLGHNPAGTSTQDLFFHWAQKGLQAYNWGSPQNMLH	316
Qy	324	YGHLPAPFADLSIPSPSLPTWCGYGLDALADVTQVTRIRELGSTPELly---ICDYGH	380
Db	317	YNQKTPPYDVDSAMTYPVIAW---NGGHDILADPDQVAMLLPKL---PNLLYHKELPYNH	371
Qy	381	IDFVMSYKAKDDVYVDLIRFLREN	404
Db	372	LDFIWMADAPQEVNVEITVMTMAED	395

RESULT 10 .

Q9DD6P3 PRELIMINARY; PRT; 395 AA.

AC Q9DD6P3;

ID Q9DD6P3;

DT 01-JUN-2001 (T-EMBLrel. 17, Created)

DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)

DE Adult male tongue cDNA, RIKEN full-length enriched library,

DE clone:2310069P19, full insert sequence.

GN GN310051B21RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RP [1]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TONGUE;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawara A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukuinishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK010139; BAB26725.1; -.

DR MGD; MGI:1914967; 2310051B21RIK.

DR InterPro; IPR000073; Abhydrolase.

DR InterPro; IPR000073; Lipase.

DR InterPro; IPR000379; Ser-estrs\_site.

DR Pfam; PF00561; abhydrolase; 1.

DR PROSITE; PS00120; LIPASE\_SER; 1.







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2002, 20:42:42 ; Search time 37 Seconds  
(without alignments)  
326.038 Million cell updates/sec

Title: US-09-699-652A-14  
Perfect score: 2194  
Sequence: 1 MAMAGHAPGALPLILLVWS.....DDVYVDLIRFLRENGWHNSY 410

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/2/1aa/PC10S\_COMB.pap.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	571	26.0	378	4	US-09-186-489-2
2	535.5	24.4	380	1	US-08-227-108-16
3	535.5	24.4	380	2	US-09-073-674-16
4	534	24.3	377	2	US-08-227-108-17
5	534	24.3	377	2	US-09-073-674-17
6	529	24.1	379	1	US-08-227-108-3
7	529	24.1	379	2	US-09-073-674-3
8	529	24.1	380	1	US-08-227-108-5
9	529	24.1	380	2	US-09-073-674-5
10	507.5	23.1	392	4	US-09-820-001-4
11	492.5	22.4	395	4	US-09-820-001-2
12	491.5	22.4	379	1	US-08-227-108-18
13	491	22.4	379	2	US-09-073-674-18
14	105	4.8	5588	4	US-09-036-987A-6
15	105	4.8	5588	4	US-09-370-700-6
16	94	4.3	634	1	US-07-779-049-3
17	94	4.3	634	1	US-08-080-240-3
18	93.5	4.3	1167	1	US-08-100-709-2
19	93.5	4.3	1167	1	US-08-176-865-2
20	93.5	4.3	1167	1	US-08-474-038-2
21	93.5	4.3	1167	2	US-08-779-046-2
22	93.5	4.3	1167	2	US-08-881-340-2
23	93	4.2	388	1	US-08-232-519-2
24	93	4.2	388	1	US-08-456-956-2
25	90.5	4.1	465	4	US-08-985-452-15
26	90	4.1	1676	4	US-08-487-283A-2
27	88	4.0	972	3	US-08-750-141A-2

28	87.5	4.0	298	4	US-09-355-166-6	Sequence 6, Appli
29	86	3.9	686	1	US-08-350-884-70	Sequence 70, Appl
30	86	3.9	686	1	US-08-709-173-70	Sequence 70, Appl
31	86	3.9	686	2	US-08-709-177-70	Sequence 70, Appl
32	86	3.9	841	1	US-08-350-884-86	Sequence 86, Appl
33	86	3.9	841	1	US-08-709-173-86	Sequence 86, Appl
34	86	3.9	841	2	US-08-709-177-86	Sequence 86, Appl
35	85	3.9	385	3	US-09-053-866-2	Sequence 2, Appli
36	85	3.9	385	4	US-09-479-130-2	Sequence 2, Appli
37	85	3.9	878	4	US-09-087-277-2	Sequence 2, Appli
38	85	3.9	878	4	US-09-658-499-2	Sequence 2, Appli
39	85	3.9	906	4	US-09-134-001C-3218	Sequence 3218, Ap
40	84.5	3.9	566	2	US-08-920-095-3	Sequence 3, Appli
41	84.5	3.9	566	5	PCT-US96-05800-3	Sequence 26, Appl
42	84.5	3.9	569	2	US-08-467-822-26	Sequence 26, Appl
43	84.5	3.9	569	4	US-08-432-697-26	Sequence 26, Appl
44	84.5	3.9	569	4	US-08-466-248-26	Sequence 26, Appl
45	84.5	3.9	1227	1	US-08-448-170-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-09-186-489-2  
; Sequence 2, Application US/09186489  
; Patent No. 6375947  
; GENERAL INFORMATION:  
; APPLICANT: Bolen, Paul L.  
; APPLICANT: Cihak, Paul L.  
; APPLICANT: Scharpf Jr., Lewis G.  
; TITLE OF INVENTION: Purified Recombinant Kid Pregastric Esterase, and  
; TITLE OF INVENTION: Processes for its Production and Use  
; FILE REFERENCE: 5499/3  
; CURRENT APPLICATION NUMBER: US/09/186.489  
; CURRENT FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Kid (Goat)  
US-09-186-489-2

Query Match	26.0%	Score 571;	DB 4;	Length 378;
Best Local Similarity	35.5%	Pred. No. 1.le-53;		
Matches 130;	Conservative 66;	Mismatches 154;	Indels 16;	Gaps 9;
QY	49	QLLLPLGYPCTEHNVEYTKDGLLSLQHIPHGKNKAAD-STGPPVFLQHLFGQGGDTWFIN	107	
Db	17	QMISFWGYPSEMHKVTADGYILQVYRIPHGKNDANHLGQRPVFLQHLGLASATNWSN	76	
QY	108	SAEQSLGVLADNGFDVWIGNVGRTRMSKGHSTFVSHDKLFWDSWQELAEYDILLAMGY	167	
Db	77	LPNNSLGLLADAGYDVWLGNSRGTWAEHLVYSPDSPEFWAFSFDENAEYDLPSTIDF	136	
QY	168	VYTVT-QSKILVGHSGQITMGLAAL-TMPEIVKMSSAALLCPISYLDHVSASFVLRVAV	225	
Db	137	ILKRTGQKLLHVGHSQGTIGFVAFSTNPTLAEKIEVHALAPVATVKTOSLFNKUAL	196	
QY	226	AMHLDQMLV---TMGIHOLNFRSDMGVQIVDSLCDEHVD---CNNLLSAITG-ENCCFNT	279	
Db	197	IPHFLFKIIFGNKMEYPHNFFEQFLGVEV---CSRETLDVLCNKALFAITCADKNFNW	252	
QY	280	SRIDYLYEYEPHPSKTNLHLFQMRKGTFAKYDYL-LGNLRRYGHLLRPPAFDLSSTP	338	
Db	253	SRLDYVAHNPAVASVQNILHWRQAISKGFQAFDQWASVENLMHYNQPTPIYNLTAMN	312	
QY	339	ESLPTWNGYGGDLADYTDVQRTRELSTPELIIYIGDYGHIDFMSVKAKDDVYVDLI	398	
Db	313	VFIAYWS--AGODLLADPDQVDLLLSKLSNLTHHKEIPNYNHLDFIWMADPAQEVYNEII	370	
QY	399	REFLEN 404		





[illegible][illegible][illegible]



Db 137 FILKGTGDKLHYVGHSGTIGTGFIAFTNPKLAKRIKTFYALAPVATVKYTETLLNKL 196  
Qy 219 ---SFVLRAVAMH-----LDQMLVTMGTHQLNFRSDMGVQIVDSLCDGEHVD--CNN 265  
Db 197 LVPSFLFLKIFGNKTFYPHFFDQFLAT-----EVCSTRTVDLLCSN 238  
Qy 266 LLSAITG-ENCCFNTSRIDYLEYEPHPSSTKNLHLMFQIRKGTFAKYDYGL-LGNLRR 323  
Db 239 ALFIICGDTNMLNSRLDVLSHNPAGTSVQNLHWSQAVKSGKFAFDGSPVQNMH 298  
Qy 324 YGHLRPPADLSSIPESLPIMWGYGGLDALADVTVDQRTIRELGSTPELly---IGDYCH 380  
Db 299 YHQSMPYPYNTDMHVPJAVN--NGNDLLADPHDVLKSL---PNLIYHRKIPPNH 353  
Qy 381 IDVMSVKAKDDVYVDLIRFL 401  
Db 354 LDFIWMADPAQAVYNEIVSMM 374  
RESULT 8  
US-08-227-108-5  
; Sequence 5, Application US/08227108  
; Patent No. 5807726  
; GENERAL INFORMATION:  
; APPLICANT: Blanchard, Claire  
; APPLICANT: Benicourt, Claude  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/227,108  
; FILING DATE: 03-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Panucci, Allan A.  
; REGISTRATION NUMBER: 30,256  
; REFERENCE/DOCKET NUMBER: 7620-033  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-227-108-5

Query Match 24.1%; Score 529; DB 1; Length 380;  
Best Local Similarity 32.5%; Pred. No. 4.4e-49;  
Matches 124; Conservative 70; Mismatches 135; Indels 52; Gaps 13;  
Qy 49 QLLPLGYPCTEHNVTGDKGFLLSLQHIPHGKNAADSTG--PPVFLQHLGFGQGDWTFI 106  
Db 19 QMITWGYPAEYEVVTEGVDYILGIDRIPYGR-KNSENIGRRPVAFLQHLGLASATNWS 77  
Qy 107 NSAEQSLGYILLADGFDVWIGNVRGTRWSKGHSTFSVHDKLFWDSWQELAEYDLLAMLG 166  
Db 78 NLPNNSLAFILADAGYDVLGNLSRGNTWARRNLXYSPODSVEFWAFSFDMAKYDLPATID 137

Qy 167 YVYTVT-OSKILYVGHSGTIGTGLAAL-TMPEIVKMISSAALLCPISYLDHWSA----- 218  
Db 138 FILKGTGDKLHYVGHSGTIGTGFIAFTNPKLAKRIKTFYALAPVATVKYTETLLNKL 197  
Qy 219 ---SFVLRAVAMH-----LDQMLVTMGTHQLNFRSDMGVQIVDSLCDGEHVD--CNN 265  
Db 198 LVPSFLFLKIFGNKTFYPHFFDQFLAT-----EVCSTRTVDLLCSN 239  
Qy 266 LLSAITG-ENCCFNTSRIDYLEYEPHPSSTKNLHLMFQIRKGTFAKYDYGL-LGNLRR 323  
Db 240 ALFIICGDTNMLNSRLDVLSHNPAGTSVQNLHWSQAVKSGKFAFDGSPVQNMH 299  
Qy 324 YGHLRPPADLSSIPESLPIMWGYGGLDALADVTVDQRTIRELGSTPELly---IGDYCH 380  
Db 300 YHQSMPYPYNTDMHVPJAVN--NGNDLLADPHDVLKSL---PNLIYHRKIPPNH 354  
Qy 381 IDVMSVKAKDDVYVDLIRFL 401  
Db 355 LDFIWMADPAQAVYNEIVSMM 375  
RESULT 9  
US-09-073-674-5  
; Sequence 5, Application US/09073674  
; Patent No. 5998189  
; GENERAL INFORMATION:  
; APPLICANT: Blanchard, Claire  
; APPLICANT: Benicourt, Claude  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Warner-Lambert Company  
; STREET: 2800 Plymouth Road  
; CITY: Ann Arbor  
; STATE: Michigan  
; COUNTRY: U.S.A.  
; ZIP: 48105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,674  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crissey, Todd M.  
; REGISTRATION NUMBER: 37,807  
; REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 734 622-7530  
; TELEFAX: 734 622-1553  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-073-674-5

Query Match 24.1%; Score 529; DB 2; Length 380;  
Best Local Similarity 32.5%; Pred. No. 4.4e-49;  
Matches 124; Conservative 70; Mismatches 135; Indels 52; Gaps 13;  
Qy 49 QLLPLGYPCTEHNVTGDKGFLLSLQHIPHGKNAADSTG--PPVFLQHLGFGQGDWTFI 106  
Db 19 QMITWGYPAEYEVVTEGVDYILGIDRIPYGR-KNSENIGRRPVAFLQHLGLASATNWS 77  
Qy 107 NSAEQSLGYILLADGFDVWIGNVRGTRWSKGHSTFSVHDKLFWDSWQELAEYDLLAMLG 166



STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,108  
FILING DATE: 03-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fanucci, Allan A.  
REGISTRATION NUMBER: 30,256  
REFERENCE/DOCKET NUMBER: 7620-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-227-108-18

Query Match 22.4%; Score 491; DB 1; Length 379;  
Best Local Similarity 31.7%; Pred. No. 6.2e-45;  
Matches 117; Conservative 78; Mismatches 152; Indels 22; Gaps 12;

QY 49 QLLPLGYPCTEHNHYETKDGFLLSLOHHPHG-KNKAADSTGPPVFLQHGFGGDTWFIN 107  
Db 18 QMISWGYPSEKYEYVETDGYILEVNRIPYKKNNGRQRPVFLQHGGLLASASNNWISN 77  
QY 108 SAEQSLGVLADNGFDVWIGNVGRTRWSKGHSTFVSVDKLFWDWSQWELAEYDILLAMLG 167  
Db 78 LPNNSLAFLADAGYGVWLGNSRGNTWSRRNLVYSPDSVEFWAFSDEMAKYDLPATIDF 137  
QY 168 VYTVT-QSKILYVGHSGQTIMGLAAL-TMPEIVKMISSAALLCPISYLDHVSASFVLRV 225  
Db 138 IVKETGQEKLVHVGHSQGTIGFIAFSTNPKLAERIKTFYALAPVATVKY-TKSLVNLK 196  
QY 226 AMHLDQMLVTMG---IHOLNFRSDMGVQIVDSLCDE--HVCNLLSAITG-ENCCFNT 279  
Db 197 FIPPTMFKIIFGDKIFYPHNFFDQF---LATQVCSRETINLVICSNALFTICGDFSANLNM 253  
QY 280 SRIDYLYEPHPSTKKNLHLLFQMRKGTFAKYDGL-LGNLRRYGHLLRPPAFDLSSTP 338  
Db 254 SRLDYVSHNPAGTSVQNMLHWTQAVKSGNFOAFNMGSPAQNVVHFNQPTPPYVNVVTAMN 313  
QY 339 ESLPTMVGGLDALADYTDVQRTIRELGSTPELly---IGDYGHIDFVMSVKAKDDYV 395  
Db 314 VPIAVMS--GGNDWLADPOVDVLLPKLSN---LIYHKEILPYNLHDFIWMANPAQEVYN 368  
QY 396 DLIRFLREN 404  
Db 369 EIIISMWAKD 377

RESULT 13  
US-09-073-674-18  
Sequence 18, Application US/09073674  
Patent No. 5998189  
GENERAL INFORMATION:  
APPLICANT: Blanchard, Claire  
APPLICANT: Benicourt, Claude  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: Recombinant Dog Gastric Lipase

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Warner-Lambert Company  
STREET: 2800 Plymouth Road  
CITY: Ann Arbor  
STATE: Michigan  
COUNTRY: U.S.A.  
ZIP: 48105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,674  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crissey, Todd M.  
REGISTRATION NUMBER: 37,807  
REFERENCE/DOCKET NUMBER: 5072-D1-66-TWC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 734 622-7530  
TELEFAX: 734 622-1553  
TELEX:  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 379 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-073-674-18

Query Match 22.4%; Score 491; DB 2; Length 379;  
Best Local Similarity 31.7%; Pred. No. 6.2e-45;  
Matches 117; Conservative 78; Mismatches 152; Indels 22; Gaps 12;

QY 49 QLLPLGYPCTEHNHYETKDGFLLSLOHHPHG-KNKAADSTGPPVFLQHGFGGDTWFIN 107  
Db 18 QMISWGYPSEKYEYVETDGYILEVNRIPYKKNNGRQRPVFLQHGGLLASASNNWISN 77  
QY 108 SAEQSLGVLADNGFDVWIGNVGRTRWSKGHSTFVSVDKLFWDWSQWELAEYDILLAMLG 167  
Db 78 LPNNSLAFLADAGYGVWLGNSRGNTWSRRNLVYSPDSVEFWAFSDEMAKYDLPATIDF 137  
QY 168 VYTVT-QSKILYVGHSGQTIMGLAAL-TMPEIVKMISSAALLCPISYLDHVSASFVLRV 225  
Db 138 IVKETGQEKLVHVGHSQGTIGFIAFSTNPKLAERIKTFYALAPVATVKY-TKSLVNLK 196  
QY 226 AMHLDQMLVTMG---IHOLNFRSDMGVQIVDSLCDE--HVCNLLSAITG-ENCCFNT 279  
Db 197 FIPPTMFKIIFGDKIFYPHNFFDQF---LATQVCSRETINLVICSNALFTICGDFSANLNM 253  
QY 280 SRIDYLYEPHPSTKKNLHLLFQMRKGTFAKYDGL-LGNLRRYGHLLRPPAFDLSSTP 338  
Db 254 SRLDYVSHNPAGTSVQNMLHWTQAVKSGNFOAFNMGSPAQNVVHFNQPTPPYVNVVTAMN 313  
QY 339 ESLPTMVGGLDALADYTDVQRTIRELGSTPELly---IGDYGHIDFVMSVKAKDDYV 395  
Db 314 VPIAVMS--GGNDWLADPOVDVLLPKLSN---LIYHKEILPYNLHDFIWMANPAQEVYN 368  
QY 396 DLIRFLREN 404  
Db 369 EIIISMWAKD 377

RESULT 14  
US-09-036-987A-6  
Sequence 6, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.

APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R.  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5588 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-036-987A-6

Query Match 4.8%; Score 105; DB 4; Length 5588;  
Best Local Similarity 21.6%; Pred. No. 0.75;  
Matches 87; Conservative 62; Mismatches 164; Indels 90; Gaps 19;

```
Qy 11 ALPLLVLVSCCGRTVSGASPA-----AALRVG-----SGSGGLCDQL 50
Db 2232 AVPL-----IVSGRTPAALSAQASALLSYLGERGDIETLDAAFSLASSRAALEER 2281
Qy 51 LPLPGYPCTEHNVTETKDFLLSLQHIPHGKNAADSTGPPV-----FLQHGGLFGGDTWF 105
Db 2282 AVVLG-----ADRETLISGLEALASGREASGVVSGSPVSGGVGFVAG--QGQWLG 2331
Qy 106 INSARQSLGYILAD-----NGFDVWIGNVRGTRWSKGHSTFVHDKLFDWDSWQELAEY 159
Db 2332 MGRGLYSVFPVADAFDEACAGLDAHLGQDVGVDRDVFSGDGLDRTLWAQSLFALQV 2391
Qy 160 DLLAMLGYYVTQSKILYVGHSGQTINGLAA---LTMPETVKMTSSAALLCPTISYLDHV 216
Db 2392 GLLSLLG--SWGVRPGVVL--GHSVGEFAAAVAGVLSLPDAARVAGRARL-----MOAL 2443
Qy 217 SASFVLRAVAMHLDQMLVTM-----GIHQLNFRSDMGVQIVDSLCDGGEHVDCNNLLSA 269
Db 2444 PSGGAMLAVAAAGEEQLRPLADRVGDGAGIAAVNAPESV-----VLSGDREVLDDIAGA 2496
Qy 270 ITGENCCFNTSRIDY-YLEYEPHPSSTKNLHLLFQMIKRGTFKAKYDY---GLLGNLRRY 324
Db 2497 LDGQGIKRRRLRVSHAFHSYRMDPM---LOEFAEIAIRSVDYRRGDLPPVSTLTGELDTA 2552
Qy 325 GHLRPPAFDLSIPESLPIMWGYGGLDALADVDVQRTIRELG 367
Db 2553 GVMATPEYVWVQVRE--PVRFA-DGVRVLAQ--QGVATIFELG 2590
```

Search completed: December 30, 2002, 20:48:12  
Job time : 42 secs

RESULT 15  
US-09-370-700-6  
Sequence 6, Application US/09370700  
Patent No. 6274350  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H  
APPLICANT: Broughton, Mary C  
APPLICANT: Crawford, Kathryn P  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Treadway, Patti J  
APPLICANT: Turner, Jan R  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/09/370,700  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/36987  
EARLIER FILING DATE: 1998-03-09  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 5588  
TYPE: PRT  
ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-6

Query Match 4.8%; Score 105; DB 4; Length 5588;  
Best Local Similarity 21.6%; Pred. No. 0.75;  
Matches 87; Conservative 62; Mismatches 164; Indels 90; Gaps 19;

```
Qy 11 ALPLLVLVSCCGRTVSGASPA-----AALRVG-----SGSGGLCDQL 50
Db 2232 AVPL-----IVSGRTPAALSAQASALLSYLGERGDIETLDAAFSLASSRAALEER 2281
Qy 51 LPLPGYPCTEHNVTETKDFLLSLQHIPHGKNAADSTGPPV-----FLQHGGLFGGDTWF 105
Db 2282 AVVLG-----ADRETLISGLEALASGREASGVVSGSPVSGGVGFVAG--QGQWLG 2331
Qy 106 INSARQSLGYILAD-----NGFDVWIGNVRGTRWSKGHSTFVHDKLFDWDSWQELAEY 159
Db 2332 MGRGLYSVFPVADAFDEACAGLDAHLGQDVGVDRDVFSGDGLDRTLWAQSLFALQV 2391
Qy 160 DLLAMLGYYVTQSKILYVGHSGQTINGLAA---LTMPETVKMTSSAALLCPTISYLDHV 216
Db 2392 GLLSLLG--SWGVRPGVVL--GHSVGEFAAAVAGVLSLPDAARVAGRARL-----MOAL 2443
Qy 217 SASFVLRAVAMHLDQMLVTM-----GIHQLNFRSDMGVQIVDSLCDGGEHVDCNNLLSA 269
Db 2444 PSGGAMLAVAAAGEEQLRPLADRVGDGAGIAAVNAPESV-----VLSGDREVLDDIAGA 2496
Qy 270 ITGENCCFNTSRIDY-YLEYEPHPSSTKNLHLLFQMIKRGTFKAKYDY---GLLGNLRRY 324
Db 2497 LDGQGIKRRRLRVSHAFHSYRMDPM---LOEFAEIAIRSVDYRRGDLPPVSTLTGELDTA 2552
Qy 325 GHLRPPAFDLSIPESLPIMWGYGGLDALADVDVQRTIRELG 367
Db 2553 GVMATPEYVWVQVRE--PVRFA-DGVRVLAQ--QGVATIFELG 2590
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw\_model

Run on: December 30, 2002, 20:43:42 ; Search time 71 Seconds  
(without alignments)  
99.983 Million cell updates/sec

Title: US-09-699-652A-14

Perfect score: 2194

Sequence: 1 MAMAGAPGALPILLVVS.....DDVYDLIRFLRENGWHSY 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 109717 seqs, 17314136 residues

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*

2: /cgn2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/1/pubpaa/PTCTUS\_PUBCOMB.pep.\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*

10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*

12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*

13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	546	24.9	398	9	US-10-042-431-75
2	531	24.2	398	9	US-10-056-744B-2
3	515.5	23.5	399	9	US-10-042-431-41
4	507.5	23.1	392	12	US-10-003-302-4
5	502	22.9	374	10	US-09-811-825-4
6	502	22.9	390	9	US-10-042-431-49
7	502	22.9	398	10	US-09-811-825-2
8	502	22.9	409	10	US-09-835-996A-21
9	502	22.9	423	9	US-10-042-431-47
10	492.5	22.4	395	12	US-10-003-302-2
11	384.5	17.5	221	9	US-10-042-431-50
12	264	12.0	127	9	US-10-056-744B-5
13	114.5	5.2	144	9	US-10-042-431-52
14	90.5	4.1	465	10	US-09-923-779-151
15	90.5	4.1	473	10	US-09-925-297-567
16	90	4.1	1602	10	US-09-778-927A-59
17	89	4.1	305	10	US-09-815-242-12337
18	88	4.0	972	10	US-09-944-807-88
19	87.5	4.0	298	10	US-09-950-368-6

20	87.5	4.0	1251	10	US-09-778-927A-58	Sequence 58, Appl
21	87	4.0	315	12	US-10-080-644-2	Sequence 2, Appli
22	85.5	3.9	297	12	US-10-080-644-7	Sequence 7, Appli
23	85.5	3.9	2799	9	US-10-151-736-4	Sequence 4, Appli
24	85	3.9	232	10	US-09-856-578-4	Sequence 4, Appli
25	84.5	3.9	569	9	US-09-895-913A-252	Sequence 252, App
26	84	3.8	1079	9	US-09-895-793-947	Sequence 947, App
27	84	3.8	1079	10	US-09-822-827-947	Sequence 947, App
28	83.5	3.8	252	10	US-09-452-239-14	Sequence 14, Appl
29	83.5	3.8	467	10	US-09-735-933-6	Sequence 6, Appli
30	82.5	3.8	676	10	US-09-815-242-13513	Sequence 13513, A
31	82	3.7	686	9	US-09-881-239-3	Sequence 3, Appli
32	82	3.7	686	10	US-09-881-654-2	Sequence 2, Appli
33	82	3.7	728	9	US-09-881-239-1	Sequence 1, Appli
34	81	3.7	632	10	US-09-929-955-29	Sequence 29, Appl
35	81	3.7	686	10	US-09-929-955-17	Sequence 17, Appl
36	81	3.7	686	10	US-09-929-955-31	Sequence 31, Appl
37	81	3.7	686	10	US-09-929-955-32	Sequence 32, Appl
38	81	3.7	686	10	US-09-929-955-43	Sequence 43, Appl
39	81	3.7	686	10	US-09-929-955-44	Sequence 44, Appl
40	81	3.7	686	10	US-09-929-955-45	Sequence 45, Appl
41	81	3.7	686	10	US-09-929-955-46	Sequence 46, Appl
42	81	3.7	686	10	US-09-929-955-47	Sequence 47, Appl
43	81	3.7	686	10	US-09-929-955-48	Sequence 48, Appl
44	81	3.7	686	10	US-09-929-955-49	Sequence 49, Appl
45	80	3.6	2894	10	US-09-941-611-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1

US-10-042-431-75

Sequence 75, Application US/10042431

Publication No. US20020182675A1

GENERAL INFORMATION:

APPLICANT: MCCARTHY, Sean A

APPLICANT: FRASER, Thomas M

APPLICANT: FRASER, Christopher C

APPLICANT: SHARP, John D

TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,

TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES

FILE REFERENCE: 10147-602

CURRENT APPLICATION NUMBER: US/10/042,431

CURRENT FILING DATE: 2001-10-25

PRIOR APPLICATION NUMBER: US 09/333,159

PRIOR FILING DATE: 1999-06-14

PRIOR APPLICATION NUMBER: US 09/578,063

PRIOR FILING DATE: 2000-05-24

NUMBER OF SEQ ID NOS: 79

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 75

LENGTH: 398

TYPE: PRT

ORGANISM: Homo sapiens

US-10-042-431-75

Query Match	24.9%	Score 546;	DB 9;	Length 398;
Best Local Similarity	33.9%;	Pred. No. 2.1e-44;		
Matches 130;	Conservative 70;	Mismatches 132;	Indels 52;	Gaps 13;
QY	49	QLLLPLGYPCTEHNVTEDKDFLLSLQHPHGKNAADSTG--PPVFLQHLFGQGDWTWFI	106	
Db	37	QMTYGPYNEEYVETEDGYILEVNRIPYGR-KNSGNTGQRPVFLQHLGLASATNWIS	95	
QY	107	NSAEOSLGVILADNDFVWIGNVGRTRMSKGHSTSVHDKLFWDSWQELAYDILLAMIG	166	
Db	96	NLPNNSLAFILADAGYDVLGNRGNTWARRNLYTSPDSVEFWAFSEDMAKYDLPATID	155	
QY	167	VYVYTVT-QSKILYVHSGQCTINGLAAL-TMPEIVKMISSAALICPISYLDHVSASF-VLR	223	
Db	156	FIVKTKGQKQLHYVGHSGCTTGTFATFSTPNPSLAKRIKTFYALAPVATVYKTSINKLR	215	





; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-003-302-4

Query Match 23.1%; Score 507.5; DB 12; Length 392;  
Best Local Similarity 32.4%; Pred. No. 1e-40;  
Matches 130; Conservative 57; Mismatches 141; Indels 73; Gaps 12;

QY 42 GSGGL-----CDQLLPLGVPCTEHNVTGKGLLSLQHIHPHG-KNKAADSTGP 89  
Db 19 GSGGLTAVDPETNMNVSEIISYWGPPSEYLVETEDGYILCLNRIPHGKKNHSDGPKP 78  
QY 90 PVFLQHLFGQGDWTFINSAESGLGVLADNGFDVWGNVGRTRSKGHSSTSVHDKLEW 149  
Db 79 VVFLQHLGLADSSNWTNLSNLSGLFILADAGFDVWGNVGRTRSKGHSSTSVHDKLEW 138  
QY 150 DMSWQELAEYDILLMLGLYVYVTVT-OSKILYVGHSGQTIMGALALT-MPEIVKMISSAALL 207  
Db 139 AFSYDEMAKYDLPASINFILNLTGQGVYVGHSGQTIMGAFIAFQIPELAKRIKMFAL 198  
QY 208 CPISYLDHVSASFVLRAVAM-----HLDQMLVTMGIHOLNFRSDMGVQIVDSLCD----- 257  
Db 199 GPVA-----SVAFCSPMAKLGRLPDHL-----IKDLFGDKLEFL 232  
QY 258 -----GEHVDCNLLSALTGENCF-----NTSRIDYLEYEPHPSSTKNL 298  
Db 233 POSAFKWLGTHTVHVLKELCG-NLCFLCGFNERNLMSRDVYVTHSPAGTSVQNM 291  
QY 299 HHLFQMRKGTFAKYDYLGLL-NLRRYGHLPAPFDLSIPESLPIMWGYGGLDALADVT 357  
Db 292 LHSQAVKFKQAFDWSGSAKYNHYQSYPTYNKDLVLPVAVWS--GGHDLADVI 349  
QY 358 DVORTIRELGSTPELLYIGDYHIDFVMSVKAKDDVYVDLI 398  
Db 350 DNVILLTQITNLVFNHESIPWEHLDFTWGLDAPWRLYNKII 390

## RESULT 5

US-09-811-825-4  
; Sequence 4, Application US/09811825  
; Patent No. US20020144297A1  
; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN LIPASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001170  
; CURRENT APPLICATION NUMBER: US/09/811,825  
; CURRENT FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 374  
; TYPE: PRT  
; ORGANISM: Human  
US-09-811-825-4

Query Match 22.9%; Score 502; DB 10; Length 374;  
Best Local Similarity 34.0%; Pred. No. 3.2e-40;  
Matches 124; Conservative 67; Mismatches 146; Indels 28; Gaps 13;

QY 55 GYPCTEHNVTGDKGFLLSLQHIHPHGKNKAADSTG--PPVFLQHLFGQGDWTFINSAES 112  
Db 8 GYPCEEYEVATEDGYILSVNRIPGLVQ-PKKTGSRPVVLLQHLGVGGASNNISLNPNS 66  
QY 113 LGYILADNGFDVWGNVGRTRSKGHSSTFSVHDKLFWDSWQELAEYDILLMLGLYVYVTV 172  
Db 67 LGFILADAGFDVWGNVGRTRSKGHSSTFSVHDKLFWDSWQELAEYDILLMLGLYVYVTV 126

QY 173 -OSKILYVGHSGQTIMGALALT-TMPEIVKMISSAALLCPISYLDHVS-----SFVLRAVAM 227  
Db 127 GOEKIYVGVYSGTGTMGFIATFSTMPELAQIKMYFALAPIATVKHAKSPCTKFL----- 181  
QY 228 HLDQMLV--TMGIHOLNFRSDMGVQIVDSLCDGEHVD--CNLLSAITGENC-CFNTSRI 282  
Db 182 -LPDMMIKGLFGKKEFLYQTRFLQRLVLYLCGVILDDQICSNIMLLLGGFTNNMMSRA 240  
QY 283 DYLYEVEPHPSSTKNLHHLFQMRKGTFAKYDYG-LGNLRRYGHLPAPFDLSSTIPESL 341  
Db 241 SYAAHTLAGTSVQNLHWSQAVNSGELRAFDMGSETKNLKCNOPTPVRYRVROM--TV 298  
QY 342 PIWMGYGGLDALADVTDVORTIRELGSTPELLY---IGDYHIDFVMSVKAKDDVYVDLI 398  
Db 299 PTAMTGGQDLSNPEDYKMLLSEV---TNLIYHKNIPEMAHVDFITWGLDAPHRMYNEII 355  
QY 399 RFLRE 403  
Db 356 HLMQ 360

## RESULT 6

US-10-042-431-49

; Sequence 49, Application US/10042431  
; Publication No. US20020182675A1  
; GENERAL INFORMATION:  
; APPLICANT: MCCARTHY, Sean A  
; APPLICANT: BARNES, Thomas M  
; APPLICANT: FRASER, Christopher C  
; APPLICANT: SHARP, John D  
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,  
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES  
; FILE REFERENCE: 10147-602  
; CURRENT APPLICATION NUMBER: US/10/042,431  
; CURRENT FILING DATE: 2001-10-25  
; PRIOR APPLICATION NUMBER: US 09/333,159  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: US 09/578,063  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-042-431-49

Query Match 22.9%; Score 502; DB 9; Length 390;  
Best Local Similarity 34.0%; Pred. No. 3.3e-40;  
Matches 124; Conservative 67; Mismatches 146; Indels 28; Gaps 13;

QY 55 GYPCTEHNVTGDKGFLLSLQHIHPHGKNKAADSTG--PPVFLQHLFGQGDWTFINSAES 112  
Db 24 GYPCEEYEVATEDGYILSVNRIPGLVQ-PKKTGSRPVVLLQHLGVGGASNNISLNPNS 82  
QY 113 LGYILADNGFDVWGNVGRTRSKGHSSTFSVHDKLFWDSWQELAEYDILLMLGLYVYVTV 172  
Db 83 LGFILADAGFDVWGNVGRTRSKGHSSTFSVHDKLFWDSWQELAEYDILLMLGLYVYVTV 142  
QY 173 -OSKILYVGHSGQTIMGALALT-TMPEIVKMISSAALLCPISYLDHVS-----SFVLRAVAM 227  
Db 143 GQEKIYVGVYSGTGTMGFIATFSTMPELAQIKMYFALAPIATVKHAKSPCTKFL----- 197  
QY 228 HLDQMLV--TMGIHOLNFRSDMGVQIVDSLCDGEHVD--CNLLSAITGENC-CFNTSRI 282  
Db 198 -LPDMMIKGLFGKKEFLYQTRFLQRLVLYLCGVILDDQICSNIMLLLGGFTNNMMSRA 256  
QY 283 DYLYEVEPHPSSTKNLHHLFQMRKGTFAKYDYG-LGNLRRYGHLPAPFDLSSTIPESL 341  
Db 257 SYAAHTLAGTSVQNLHWSQAVNSGELRAFDMGSETKNLKCNOPTPVRYRVROM--TV 314  
QY 342 PIWMGYGGLDALADVTDVORTIRELGSTPELLY---IGDYHIDFVMSVKAKDDVYVDLI 398



Qy 226 AMHLDQMLVTMGIHQINFRSDMGVQIYDS-LCDGE--HYDCNNLLISAITGEN-CCFNITS 281

Dd 227 -----FCTKGF-----FLEDKTKTIASTKICNNKILWLICSEFMSLWAGSNKKNNMQSR 275

Qy	282	IDYYLEYEPHPS	TKNLJHHLFQMI	KRGTFAKYDYL	-LGNLRRYGHULPPAFDLSSTPES	340
		:	:	:	:	:
		:	:	:	:	:
		:	:	:	:	:
Db	276	MDVYMSHAPT	GSSVHNLIHKQLYH	SDEFRAYDGNK	HYNQSHPTIYDLTAMKVP	338
		:	:	:	:	:
		:	:	:	:	:
		:	:	:	:	:
Qy	341	LPIWMGVGGDLA	ADVTVDVORTTREL	GSTPELLYIGDYGH	--IDFVMSVRKADVVYVDLI	399
		:	:	:	:	:
		:	:	:	:	:
		:	:	:	:	:
Db	336	TAIWA--GGHDV	LCTPDQDVARILPQ	IKSLSLVLSLPEWE	PTDFVWGLDAPQRMF----	388
		:	:	:	:	:
		:	:	:	:	:
		:	:	:	:	:
Qy	399	REFRENCWHN	408			
		:	:			
		:	:			
Db	390	-----SGNHN	394			
		:	:			
		:	:			
		:	:			

RESULT 11  
US-110-042-431-50  
; Sequence 50, Application US/10042431  
; Publication No. US20020182675A1  
; GENERAL INFORMATION:  
; APPLICANT: MCCARTHY, Sean A  
; APPLICANT: BARNES, Thomas M  
; APPLICANT: FRASER, Christopher C  
; APPLICANT: SHARP, John D  
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC.

Qy 399 RELRE 403  
:::  
Db 405 HLMCO 409

```
; PRIOR FILING DATE: 1998-06-14  
; PRIOR APPLICATION NUMBER: US-09/578,063  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 50  
; LENGTH: 221  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
US-10-042-431-50
```

Query Match      17.5%; Score 384.5; DB 9; Length 221;  
Best Local Similarity    46.4%; Pred No. 2.6e-29;  
Matches         78; Conservative    24; Mismatches    61; Indels            5; Gaps

Oy    55    GYPCETHVETKDGFLSLQIHPHGKNKAADSTG--PPVFLOHGLFGGGDTWTFINSAEQS    112  
        ||||| : | : ||| : || : | : | : ||||| : | : | : |  
Db    24    GYPCEEYEVAATEDGYILSVNRIPRGLVO-PKKTGSRPVVLLQHQLGVGGASNWSLNPNNNS    82

Vv    113    ICGYLINDGNEDVTICNVCEBSCFCHESVDPKIENDSQSOTAFVNLIAMIQAVNYTVT    179

Db	83	LGFLTAGEDVWNGSRGNWSRKHKTLSDQDEFWAFSYDENAREDLPAVINFLQKT
Qy	173	-QSKILYGVHSQGTIMGALAL-TNPEIVKVMISSAALICPTSYLDHVA 218
Db	143	GOEXIYYGVGSQGTMGFIASFSTMPELAQKTKMYFALAPATVTRHAKS 190

```
RESULT 12
US-10-056-744B-5
; Sequence 5, Application US/10056744B
; Publication No. US20020193303A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 58860, A HUMAN CHOLESTERYL ESTER
; FILE REFERENCE: MPI2001-026P1RNM
; CURRENT APPLICATION NUMBER: US/10/056.744B
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 60/264,167
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-056-744B-5

Query Match          12.08; Score 264; DB 9; Length 127;
Best Local Similarity 44.28; Pred. No. 3.9e-16;
Matches 53; Conservative 18; Mismatches 47; Indels 2; Gaps 2;

Qy 104 WFINSAEQSLGYILLADNGFDVWIGNVGRTRWSKGHSTFSVHDKLFWDMWQELAEYDILLA 163
Db 1 WIANGPNKSLAFILADAGYVWLGNSRGNTYSRKHKVLNPSHSEFWDFSWHEMGMYDLPA 60

Qy 164 MGYVYVVT-OSKILYVGHSGGTINGLAALT-MPEIVKMISAAALLCPISYLDHVSASFV 221
Db 61 MIDYVLETTGOEKLHYVGHSGQTTFFVWMLSERPEYNEKIKSFHALAPVAYMKHVRSPV 120

RESULT 13
US-10-042-431-52
; Sequence 52, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042.431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 52
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-52

Query Match          5.28; Score 114.5; DB 9; Length 144;
Best Local Similarity 24.68; Pred. No. 0.00089;
Matches 32; Conservative 30; Mismatches 59; Indels 9; Gaps 4;

Qy 278 NTSRIDYILEYEPHPSSTKNLHLFQMIRKGTFAKYDYL-LGNLRRYGHLPAPAFDLSS 336
Db 6 NMSRSVYAAHTLAGTSVQNTLHWSQAVNSGELRFDWGSSTKLEKCNQPTPVRYVRD 65

Qy 337 IPESLPIWMG-GDADADVTQIRELGSTPELly---IGDYGHIDFVMSYKAKDDV 393
Db 66 M--TVPTAMWTGGQDLNSNPEDVKMLSEV---TNLIYHKNIPEWAHVDFTWGLDAPHRM 120
```

```
Qy 394 YVDLIRFLRE 403
Db 121 YNEIHLMOQ 130

RESULT 14
US-09-923-779-151
; Sequence 151, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyte, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923.779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-779-151

Query Match          4.18; Score 90.5; DB 10; Length 465;
Best Local Similarity 20.48; Pred. No. 0.89;
Matches 80; Conservative 55; Mismatches 146; Indels 111; Gaps 22;

Qy 57 PCTPHNVETKDFLLSLQHIPHG-KNKAADSTG-----PPVFLQHGFLF-QGGDTW 104
Db 45 PWSPKDVNTR--FLLYTNENPNFQEAADSSISGNSFKTNKTRTFIHGFIKGEENW 102

Qy 105 FINSAE-----QSLGYILLADNGFDVWIGNVGRTRWSKGHSTFSVHDKLFWDMWQELAEY 159
Db 103 LANVCKNLKFKVESYNICVD-----WKGSR-----TGYTQASQNIQIV----GAEVAYF 148

Qy 160 DLLAMLGYVYVVTQSKILYVGHs-----QGTINGLAALT-----MPEIV 198
Db 149 --VEFLQSAFGYSFSPNVHIVGHSLGAHAAGAGRTNGTIGRTGLDPAEPCFGTPELV 206

Qy 199 KMISSAALLCPISYLDHV-----SASFVLRAVAMHLDQMLVTMGHQLNFRSDMGVQIYDVS 254
Db 207 RLDPESDAKFVDVIHTDGAPIVPLNGLFGMSQVVGHLDFPENGVMPCCKKNILSQIYD- 264

Qy 255 LCDG-----EHVDCNLLS-----AITGENC-----CF-----NTSR 281
Db 265 -IDGIWEGTRDFAACNHLRSKYKYYTDSIVNPDGFAGFPFCASYNVFTANKCFPCSPGGCPQ 323

Qy 282 IDYYLEYEPHPSSTKNLHLFQMIR--KGTFKAYDYGL---LGNLRRYGHLPAPAFDLSS 336
Db 324 MGHYA--DRYEGKTNVQCKEYFLOTGDASNFAWRKYKVSVTLSGCKKVTGCHILVSLFGNKG 381

Qy 337 IPESLPIWMG-----YGGDLADADVTQVQ 360
Db 382 NSQVETPKGTCLKPDSTHSNEFSDVDVGDQLQ 413

RESULT 15
US-09-925-297-567
; Sequence 567, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925.297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

Tue Dec 31 09:21:28 2002

; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 567
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-567

Query Match 4.1%; Score 90.5; DB 10; Length 473;  
Best Local Similarity 20.4%; Pred. No. 0.91;  
Matches 80; Conservative 55; Mismatches 146; Indels 111; Gaps 22;

QY	57	PCTEHNVETKDFLLSLQHIPHG-KNKAADSTG-----PPVFLQHLGF-QGGDTW	104
DB	53	PWSPKDVNTR--FLLYTNEPNFQEVAAADSSISGSNEKTNRKTRFIIHGFIIDRGEENW	110
QY	105	FINSAE-----OSLGYILADNGFDVWIGNVRCGRWSKGHSTFSVHDKLPWDWSQELAEY	159
DB	111	LANVCKNLKFEVSNVICVD-----WKGGR-----TGYTQASQNIQIV-----GAEVAYF	156
QY	160	DLMLAGYVYVVTQSKILYVGHs-----QGTIMGLAALT-----MPEIV	198
DB	157	--VEFLOSAGYSPSNVHVIGHSLGAHAAGEAGRTNGTIGRITGLDPAEPCFQGTPELV	214
QY	199	KMISSAALLCPISYLDHV-----SASFVLRAVAMHLDQMLVTMGIHQLNFRSDMGVQIVDS	254
DB	215	RLDPSDAKFVDVIHTDGAPIVNLGFGMSQVVGHLDPFPNGGVEMPGCKKNILSQIVD-	272
QY	255	LCDG-----EHVDCNNLLS-----AITGENC-----CF-----NTSR	281
DB	273	-IDGIWEGTRDFAACNHLRSYKYTDSIVNPDGAFGFPACASYNVFTANKCFPCPSGGCPQ	331
QY	282	IDYLEYEPHPSSTKNLHHLFOMIR--KGTFAKYDYL---LGNLRRYGHLPAPAFDLSS	336
DB	332	MGHYA--DRYPCKTNDVGOKFYLDTGDSNFAFWRYKYSVTLSGKKVGTGHIILVSLFGNKG	389
QY	337	IPESLPIMNG-----YGGLDALADVTDVQ	360
DB	390	NSKQYEIEFKGTLKPDSTHSNEFDSVDVGDLDQ	421

Search completed: December 30, 2002, 20:49:32  
Job time : 71 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2002, 17:12:16 ; Search time 3834 Seconds  
(without alignments)  
11257.032 Million cell updates/sec

Title: US-09-699-652A-13  
Perfect score: 1483  
Sequence: 1 gcacgagtcacagcgcgcc.....ctcttcaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.un.\*
- 28: em.vi.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	136.6	9.2	57564	8	AC006298	AC006298 Arabidops
2	100.8	6.8	793	8	AY048236	AY048236 Arabidops
3	94.2	6.4	2481	6	AX411234	AX411234 Sequence
4	94.2	6.4	2481	9	HSU08464	U08464 Human lysos
5	94.2	6.4	2481	9	HSU08464	U08464 Human lysos
6	94.2	6.4	2493	9	BC012287	M74775 Human lysos
7	94.2	6.4	2586	9	BC012287	BC012287 Homo sapi
8	94.2	6.4	2626	9	HSLAL	X76488 H.sapiens m
9	92.6	6.0	2982	9	AK091558	AK091558 Homo sapi
10	89.4	6.0	1272	6	AX477670	AX477670 Sequence
11	88	5.9	1417	4	BOVME	L26319 Bos taurus
12	88	5.9	2577	10	BC031933	BC031933 Mus muscu
13	86.6	5.8	618	8	AY091711	AY091711 Arabidops
14	86.4	5.8	1228	6	AX392821	AX392821 Sequence
15	85.8	5.8	41805	3	CBRG44C02	AC084610 Caenorhab
16	83.8	5.7	417	6	AX150702	AX150702 Sequence
17	83.8	5.7	417	6	AX150702	AX150702 Sequence
18	80.4	5.4	2358	10	MMLYACLY	Z31689 M.musculus
19	80	5.4	1365	9	HSGLR	X05997 Human mRNA
20	80	5.4	1367	6	A01046	A01046 H.sapiens m
21	80	5.4	1367	6	A12714	A12714 Pregastric
22	78.6	5.3	41014	3	CEF01G10	Z81055 Caenorhabd
23	78.4	5.3	1336	6	A01157	A01157 R.norvegicu
24	78.4	5.3	1355	10	RNLIP	X02309 Rat mRNA fo
25	78.4	5.3	1378	6	A26689	A26689 Precursor o
26	78.4	5.3	1378	6	A26690	A26690 Precursor o
27	77.6	5.2	1211	3	DMLIPASE3	Y14367 Drosophila
28	77.6	5.2	25416	2	AC013100	AC013100 Drosophill
29	77.6	5.2	82393	3	AC003923	AC003923 Drosophill
30	77.6	5.2	194672	3	AC008360	AC008360 Drosophill
31	77.6	5.2	225668	3	AE003699	AE003699 Drosophill
32	72.8	4.9	2506	9	AK096406	AK096406 Homo sapi
33	71.8	4.8	16173	3	CEF54F3	Z79696 Caenorhabd
34	71	4.8	4543	2	AC017119	AC017119 Drosophill
35	71	4.8	96018	3	AE003783	AE003783 Drosophill
36	71	4.8	172748	3	AC018484	AC018484 Drosophill
37	71	4.8	186695	2	AC013351	AC013351 Drosophill
38	70.6	4.8	23525	2	AC116988	AC116988 Dictyoste
39	69.6	4.7	40114	3	AF067942	AF067942 Caenorhab
40	68.6	4.6	1473	6	AX392824	AX392824 Sequence
41	68	4.6	3144	10	S81497	S81497 lysosomal a
42	67.2	4.5	1185	6	AX472752	AX472752 Sequence
43	66.2	4.5	93932	3	AC004766	AC004766 Drosophill
44	66.2	4.5	175007	3	AC008342	AC008342 Drosophill
45	66.2	4.5	211784	2	AC017903	AC017903 Drosophill

ALIGNMENTS

RESULT 1  
AC006298/c  
LOCUS  
DEFINITION Arabidopsis thaliana chromosome 2 clone F15A23 map mi398, complete sequence.  
ACCESSION AC006298  
VERSION AC006298.13 GI:20197742  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
REFERENCE 1 (bases 1 to 57564)  
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,

AC006298 57564 bp DNA linear PLN 27-FEB-2002  
Arabidopsis thaliana chromosome 2 clone F15A23 map mi398, complete sequence.





```
repeat_region      ARNOLD3."
13326..13525
/rpt_family="rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
ARNOLD1."
13327..13527
/rpt_family="rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
ARNOLD4."
13327..13525
/rpt_family="rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
ARNOLD3."
13410..13525
/rpt_family="rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
ARNOLD1."
13411..13525
/rpt_family="rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
ARNOLD3."
13966..14153
/rpt_family="rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
ARNOLD2."
14080..14197
/rpt_family="rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
ARNOLD4."
14080..14153
/rpt_family="rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
ARNOLD3."
14090..14197
/rpt_family="rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
ARNOLD4."
14096..14153
/rpt_family="rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
ARNOLD4."
14122..14197
/rpt_family="rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
ARNOLD1."
14124..14190
/rpt_family="rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
ARNOLD1."
14140..14290
```

Query Match 9.2%; Score 136.6; DB 8; Length 57564;

Best Local Similarity 63.0%; Pred. No. 2.le-19;

Matches 211; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 933 TTTTTCAGATGATCAGGAAGCACTTTTCGCAAGTATGACTATGGTTATTGGGAACC 992

Db 5523 TTACTCCAGTGATTCGGAAGGAACCTTTGCAATACCACTACCGGTATTTTCAAAAATC 5464

QY 993 TAAGCGCTACGGTATTCGCTCCTCCGCAATTTGACCTAAGCAGCATPACAGAAATCAC 1052

Db 5463 TACGGACTTATGGCTGTGCAAACTCCAGAAATTCATACCAACCATCCCGGCATCAT 5404

QY 1053 TGCCCATATGGATGGATATGAGGTCTTGTATGCTATGGCTGATTAACCGATGTTCCAG 1112

Db 5403 TACCAATGTGGATGGGTATGTTGGAACCTAGTGTGTAGCAGATGTGACAGATGTGGAAC 5344

QY 1113 GTACTATCAGACGTGGGATCTACACCAAGACTTCTGTACATTTGTGACTATGCCATA 1172

Db 5343 ATACTCTCCGGAACACTACCTCCAGTCCAGATTTACTATATCTTGAGGATTTATGTCACA 5284

QY 1173 TTGATTTTGTATGACGCTGAAGCGGAAGATGATGTTTATGTGGACCTTAATAAGATTTTC 1232

Db 5283 TTGACTTTTGTGCTTGGCTCAAGTGTAAAGAGATGCTCTATAAGCACATGATTCATTTT 5224

QY 1233 TTAGGAAATGGATGGCATAATAGCTATTAGGAT 1267

Db 5223 TCAGACCAAGGTTAAGTCTAGTAGTTAGTTGTTGTAATAAT 5189

RESULT 2  
AY048236

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AY048236  
Arabidopsis thaliana AT5g14180/MUA22\_18 mRNA linear PLN 05-AUG-2001  
AY048236  
AY048236.1 GI:15081687  
FLI\_CDNA.  
Arabidopsis thaliana.  
Arabidopsis thaliana  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 793)  
Koesema,E., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Shinn,P.,  
Banh,J., Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A.,  
Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T.,  
Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J.,  
Liu,S.X., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S.,  
Palm,C.J., Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,  
Southwick,A., Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S.,  
Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.  
Arabidopsis cDNA clones

Unpublished  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

2 (bases 1 to 793)  
Koesema,E., Chen,H., Cheuk,R., Kim,C.J., Meyers,M.C., Shinn,P.,  
Banh,J., Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A.,  
Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T.,  
Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J.,  
Liu,S.X., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S.,  
Palm,C.J., Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,  
Southwick,A., Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S.,  
Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.  
Direct Submission

TITLE  
JOURNAL

Submitted (09-JUL-2001) Salk Institute Genomic Analysis Laboratory  
(SIGnAL), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,  
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Koesema,E., Chen,H.,  
Cheuk,R., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J., Bowser,L.,  
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G.,  
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,  
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,  
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,  
Davis,R.W., Theologis,A., and Ecker,J.R.

Koesema,E. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally  
to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)  
contributed equally to this work as PIs.

FEATURES  
source

Location/Qualifiers  
1..793  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL05-14-M10(R13071)"  
/note="ecotype: Columbia"  
72..689  
/note="unknown protein"  
/codon\_start=1  
/product="AT5g14180/MUA22\_18"  
/protein\_id="AAK82499.1"  
/db\_xref="GI:15081688"  
/translation="MLSPVAYLSHMTTVIGDIAARTFLAEATISLGNPEFPKSLGV  
DFIKAIICLAGSDICVDLVITGKNCLNASTIDFLANEPQSTSTKNHIAOTVRD  
KELRYKNGSSDRNKKHYGOAIPPAYNISALPHELPLFFSYGLDLSADVKDFELL  
QFKYHIDKMNQVFKDYAHADFMGTAKDVVYNQVATFFRQA"

5'UTR  
CDS

3'UTR  
BASE COUNT

228 a 170 c 166 g 229 t

## ORIGIN

Query Match 6.8%; Score 100.8; DB 8; Length 793;  
Best Local Similarity 48.9%; Pred. No. 8.1e-12;  
Matches 341; Conservative 0; Mismatches 342; Indels 15; Gaps 2;

QY 578 CAGGGAACATATATGGTTTGGCGGCTTAGCATGCGCCGAAATAGTAAATAATGATTAGC 637  
DB 3 CAGGGGACGCTTAATAGGATTGCTTCAGAAAAGGGTGGTGATCAAGTGAGA 62  
QY 638 TCTGCAGCACTTCTTTGGCTATTTCTATCTGTATGATCTTTAGTGTAGTTTCTTC 697  
DB 63 TCGGGGGCGATGTTGAGTCCCGCTTCTATCTTCAGCCACATGACCAACCGTCATCGCGAT 122  
QY 698 AGAGCAGTCGCCATGCATCTTGCATCAGATGCTTTACTATGGAATTCACCAAGTGAAC 757  
DB 123 ATCGCGGCAAAACCTTCTCCCGAGGCCACTTATCTTTGGATGGCCAGAGTTTAAC 182  
QY 758 TTCCGTAGCGACATGGGGTTCCAAATAGTAGATTCTTTGTGCGATGGTGAACACGTGGAT 817  
DB 183 CCGAAAAGTGGATTAGTAGGGGATTTTATAAGGCTATATGCTTTAAAGCGGGATCGAC 242  
QY 818 TGCACAAATTTCTATCTGCGATTCACGGGAAACTGTGCTTCAATACATCAAGGATT 877  
DB 243 TGTATGATTAGTCTCTGTTATCACTGGAAGAATTGCTGCTTTAAGCGCATCAACCAT 302  
QY 878 GATTATTATTTGGAGTATGAACCTCATCATCATCGACAAATAAATCTGCACCATCTTTT 937  
DB 303 GATCTGTTCCITGGCAAGCACTCAATCTACTTCCCAAGAAGCATGATCCACCTTGCT 362  
QY 938 CAGATGATCAGGAAAGGCATTTCCCAAGATGACTATGGTTATTGG---GAAACCTA 994  
DB 363 CAACGGTAAGAGACAGGAGTTGAGAAATACACTATGGAAGCAGTATGCGCAACATA 422  
QY 995 AGGCGTACGGTCAATTTGGCTCTCCCGATTTGACCTTAAGCAGCATACCAAGATCACTG 1054  
DB 423 AAACATATCGGTTCAAGCATACACCGGCATACACATATCGCGCATCCACACAGGCT 482  
QY 1055 CCATATGATGGGATATGGAGGCTCTGATGATGCTGATGATGATGATGATGATGATGAT 1114  
DB 483 CGCTTTTTTTCAGTACGCTGGTGTATGATGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 542  
QY 1115 ACTATCAGAGCTGGGATC-----TACACCAAGCACTTCTGTACATTTGGTGAC 1162  
DB 543 CTCCTTGACCACTTTAAGTATCAGCACATCGATAGATGAATGTCAGTTTGTGAAAGAC 602  
QY 1163 TATGGCCATATGATTTGTTATGACGCTGAAGGGAAGATGATGTTATGTGGACCTA 1222  
DB 603 TATGCTCATGCTGATTTTCATCATGTTGTTTACTGCAAGAGCTGTTTATATAACCAAGTT 662  
QY 1223 ATAAGATTTCTTAGGGAATGGATGGCATAATAGCTA 1260  
DB 663 GTACTCTTTTCAAGGACAAGCTTGATAATATTATTA 700

RESULT 3  
AX411234 AX411234 2481 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 3881 from Patent WO0229103.  
DEFINITION AX411234  
ACCESSION AX411234  
VERSION AX411234.1 GI:21443939  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 3881 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES location/Qualifiers  
source 1. .2481

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. Z31690"  
BASE COUNT 692 a 477 c 513 g 799 t  
ORIGIN

Query Match 6.4%; Score 94.2; DB 6; Length 2481;  
Best Local Similarity 55.2%; Pred. No. 2.9e-10;  
Matches 249; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 194 GGCTACCCCTCACCGACGACACACGTTGAAACAAAGATGGATTCCTTTTATCTCTTCAG 253  
DB 166 GGATTCCCTAGTAGGAATACCTAGTTTGAGACAGAAGATGATATATATCTGTGCTTAAC 225  
QY 254 CATATCCACATGGCAAAATAAAGCCACAGATAGTACTGCGCCCTCCAGTT---TTTCTT 310  
DB 226 GGAATTCCTCATGGGAGGAAGAACCAATCTGCAAGGTCCCAACACAGTTGTCTTCCTG 285  
QY 311 CAACATGCTCTTTTCAGGGAGGAGACACATGGTTTCATAAACTCTGCTGAGCAATCACTT 370  
DB 286 CAACATGGCTTGTGGCAGATTCTAGTAACCTGGGTCAACAACCTTGCCACAGAGCCTG 345  
QY 371 GGGTATATCTTGTGTATACGGTTTGTGATGTTGGATGGGAATGTCCGTGGACGCGT 430  
DB 346 GGCTTCATCTTGTGTATGCTGTTTTCACGCTGGATGGCAACAGCAGAGGAATAACC 405  
QY 431 TGGAGTAAAGTCATTCACACCTTTTCTGTTGATGATAGCTTTTCTGGGATGGAGCTGG 490  
DB 406 TGGTCTCGGAAACATAAGACACATCTCAGTTTCTCAGGATGAATTTCTGGGCTTTTCAGTTAT 465  
QY 491 CAAGAGTTAGCTGAATATGACTTTTAG---CAATGCTAGGCTATGCTATACAGTCACA 547  
DB 466 GATGAGATGGCAAAATATGACCTACCAGCTTCCTCAATTCATCTGAATAAAACTGGC 525  
QY 548 CAGTCCAAATTTCTATATGTGGGCATTCACAGGGAACATAATAGGTTTGGCGCTTTG 607  
DB 526 CAAGAACAGTATATATGTGGTCAATTCACAGCACCACCTATAGGTTTATATGCAATTT 585  
QY 608 AC---GATGCCCGGAATAGTAAATGATTA 635  
DB 586 TCACAGATCCCTGAGCTGGCTAAAGGATTA 616

RESULT 4  
HSU08464 HSU08464 2481 bp mRNA linear PRI 23-JUN-1994  
LOCUS Human lysosomal acid lipase mRNA, complete cds.  
DEFINITION U08464  
ACCESSION U08464  
VERSION U08464.1 GI:505052  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
TITLE Du,H. and Gregory,G.A.  
JOURNAL Structural Conservation of Putative Functional Motifs between Mouse  
REFERENCE and Human Lysosomal Acid Lipase  
AUTHORS Unpublished  
TITLE Du,H.  
JOURNAL Direct Submission  
REFERENCE (bases 1 to 2481)  
AUTHORS

FEATURES  
source 1. .2481  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="clone hLAL"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/tissue\_type="liver"

RESULT 5	
HSLYACLY	
LOCUS	2481 bp mRNA linear PRI 17-FEB-1997
DEFINITION	H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase.
ACCESSION	Z31650
VERSION	Z31650.1 GI:506430
KEYWORDS	LAL; lipase; lysosomal acid lipase.
SOURCE	Homo sapiens.





Query Match 6.4%; Score 94.2; DB 9; Length 2626;  
Best Local Similarity 55.2%; Pred. No. 2.9e-10;  
Matches 249; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 194 GGCTACCCCTGCACGAGCAGACACGTTGAAACAAGATGATTCCTTTTATCTCTTCAG 253  
Db 278 GGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATATATCTGTGCCTTAAC 337

QY 254 CATATCCACATGGCAAAATAAAGCAGACAGATAGTACTGGCCCTCCAGTT---TTTCTT 310  
Db 338 CGAATTCCTCATGGGAGGAAGAACCATCTGACAAAGTCCCAAAACAGTTCTGTCTCCTG 397

QY 311 CAACATGGTCTTTTTCAGGAGGAGACACATGGTTCAATAAGCTCTGCTGAGCAATCACTT 370  
Db 398 CAACATGGTCTGCTGCGACATCTAGTACTGGTTCACAAACCTTGCCACAGCAGCCTG 457

QY 371 GGTATATCCTTGTGATACAGCGTTTGTGATTTGGATTTGGAATGTCGTTGGAACGGT 430  
Db 458 GGTCTCATCTCTGTGATGCTGCTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 517

QY 431 TGGAGTAAAGGTCATTCACCTTTTCTGTTTCATGATAGCTTTTCTGGGATTTGGAGCTGG 490  
Db 518 TGTCTCGGAACAATGACACTCTAGTTTCTCAGGATGAATCTGGGCTTTCAGTTAT 577

QY 491 CAAGATTTAGCTGAATATGACCTTTTAG---CAATGCTAGGCTATGTTATACAGTCACA 547  
Db 578 GATGAGATGGCAAAATATGACCTACAGCTTCCATTAACCTTCAATCTGAATAAATGCG 637

QY 548 CAGTCCAAATCTATATGTTGGGCAATTCACAGGGAATATATGTTGGTTGGCGCTTTG 607  
Db 638 CAAGAACAAAGTATATATGTTGGGCTCAATCTCAAGGCACCACTATAGTTTATAGCAAT 697

QY 608 AC---GATCCCGAAATAGTAAATGATTA 635  
Db 698 TCACAGATCCCTGAGCTGCTAAAGGATTA 728

RESULT 9  
AK091558  
LOCUS Homo sapiens cDNA FLJ34239 fls, clone FCBBF3027755, highly similar  
to LYSOXOMAL ACID LIPASE/CHOLESTERYL ESTER HYDROLASE PRECURSOR (EC  
3.1.1.13).  
AK091558  
VERSION oligo capping; fis (full insert sequence).  
KEYWORDS Homo sapiens fetal brain cDNA to mRNA, clone\_lib:FCBBF3  
SOURCE clone:FCBBF3027755.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,  
Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,  
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,  
Yamashita,H., Chiba,Y., Suigiyama,T., Irie,R., Otsuki,T., Sato,H.,  
Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K.,  
Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M.,  
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,  
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project

TITLE Unpublished  
JOURNAL 2 (bases 1 to 2982)  
REFERENCE Isogai,T. and Yamamoto,J.  
AUTHORS Direct Submission  
TITLE Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7  
JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.

#### FEATURES

source  
1. .2982  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="FCBBF3027755"  
/tissue\_type="brain"  
/clone\_lib="FCBBF3"  
/dev\_stage="fetal"  
/note="cloning vector: pME18SFL3"

BASE COUNT 847 a 596 c 615 g 924 t  
ORIGIN

Query Match 6.2%; Score 92.6; DB 9; Length 2982;  
Best Local Similarity 55.0%; Pred. No. 6.8e-10;  
Matches 248; Conservative 0; Mismatches 194; Indels 9; Gaps 3;

QY 194 GGCTACCCCTGCACGAGCAGACACGTTGAAACAAGATGATTCCTTTTATCTCTTCAG 253  
Db 810 GGATTCCCTAGTGAGGAATACCTAGTTGAGACAGAAGATGGATATATCTGTGCCTTAAC 869

QY 254 CATATCCACATGGCCAAATAAAGCAGCAGATAGTACTGGCCCTCCAGTT---TTTCTT 310  
Db 870 CGAATTCCTCATGGGAGGAAGAACCATCTGACAAGGTCCCAAAACAGTTGCTCTTCCTG 929

QY 311 CAACATGGTCTTTTTCAGGAGGAGACACATGGTTTCATAAACTCTGCTGAGCAATCACTT 370  
Db 930 CAACATGGCTTCTGCTGGCAGATTTCTAGTAACCTGGGTCACAAAACCTTGCCACGAGCAGCCTG 989

QY 371 GGTATATCCTTGTGATACAGCGTTTTCATTTTGGATTTGGAATGTCGTTGGAACGGT 430  
Db 990 GGTCTCATCTTCTGCTGGTGTGTTTTCACGCTGGATGGGCAACAGGAGGAATACC 1049

QY 431 TGGAGTAAAGGTCATTCACACCTTTTCTGTTTCATGATAAGCTTTCTCTGGGATTTGGAGCTGG 490  
Db 1050 TGGTCTCGAAACATCAGACACTCTCAGTTCTCAGGATGAATCTGGGCTTTCAGTTAT 1109

QY 491 CAAGATTTAGCTGAATATGACCTTTTAG---CAATGCTAGGCTATGTTATACAGTCACA 547  
Db 1110 GATGAGATGGCAAAATATGACCTACACCTTCCATTAACCTTCAATCTGAATAAATGCG 1169

QY 548 CAGTCCAAATCTATATGTTGGGCAATTCACAGGGAATATATGTTGGTTGGCGCTTTG 607  
Db 1170 CAAGAACAAAGTATGTTATGTTGGTCAATCTCAAGGCACCACTATAGTTTATAGCAATTT 1229

QY 608 AC---GATCCCGAAATAGTAAATGATTA 635  
Db 1230 TCACAGATCCCTGAGCTGCTAAAGGATTA 1260

RESULT 10  
AX477670  
LOCUS AX477670 1272 bp DNA linear PAT 12-AUG-2002  
DEFINITION Sequence 8 from Patent WO0246418.  
ACCESSION AX477670  
VERSION AX477670.1 GI:22216841  
KEYWORDS human.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

1  
AUTHORS Griffen,J.A., Gandhi,A.R., Ramkumar,J., Tang,Y.T., Ding,L., Yue,H.,  
Gietzen,K.J., Sapperstein,S.K., Honchell,C.D., Bruns,C.M.,  
Duggan,B.M., Xu,Y. and Lee,S.

#### TITLE

Lipid-associated molecules  
Patent: WO 0246418-A 8 13-JUN-2002;  
Incyte Genomics, Inc. (US)

#### JOURNAL

Location/Qualifiers

#### FEATURES

```
source
1..1272
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7483978CB1"
BASE COUNT      351 a      263 c      311 g      347 t
ORIGIN

Query Match      6.0%; Score 89.4; DB 6; Length 1272;
Best Local Similarity 53.6%; Pred. No. 3e-09;
Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;

QY 183 TCCTGCCACTCGGCTACCCCTGCACGACACACACGTTGAAACAAAGATGATTCCTTT 242
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 137 TCAACATCAAGCTATCCCTCTGAGGAATATCAAGTCGCAACTGAAGATGGGTATATCC 196
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 243 TATCTCTTCAGCATATCCACATGCGCAAAATAAAGCAGCAGATA---GTACTGGCCCTC 299
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 197 TTTCTGTTAACAGGATTCCTCGAGGCTAGTGCACCTAAGACAGAGTTTCAGGCCCTG 256
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 300 CAGTTTTCTTCAACATGTCCTTTTCAGGGAGGAGACACATGGTTTCATAAACTCTGCTG 359
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 257 TGTGTTACTGCAGCATGCGCTAGTTGGAGGTGCTAGCAACTGGATTTCCAACTGCCCA 316
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 360 AGCAATCACTTGGGTATATCCCTGCTGATAACGGTTTTCATGTTGGATGGGAATGTCC 419
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 317 ACAATAGCCTGGGCTTCATCTCGCAGATGCTGGTTTTCACGCTGGATGGGAACAGCA 376
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 420 GTGGAAAGCGTTGGAGTAAGGTATTCACCTTTCTCTGTTTCATGATAAGCTTTTCTGGG 479
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 377 GGGAAACCCCTGGTCTCGAAACACAGACACTCTCCATAGACCAGATGAGTCTGGG 436
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 480 ATTGGAGCTGGCAAGAGTAGTGAATATGACCTTTTAGCAATGCTAGGCT---ATGTGT 536
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 437 CTTTCAGTTATGATGAGATGGCTAGTTTGGACCTTCCTCAGTGATAAACTTTATTTCG 496
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 537 ATACACTCACAGTCCAAATCTATATGTGGGGCATTCACAGGGAACATAAATGGGTT 596
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 497 AGAAAACGGCCGAGGAAAGACTATTATGTGGCTATTTCACAGGGCACCACCATGGCT 556
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 597 TGGCGGCTTTGACGA 611
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 557 TTATTGCATTTTCCA 571
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |

RESULT 11
BOVME
LOCUS      BOVME
DEFINITION Bos taurus pregastric esterase mRNA, complete cds.
ACCESSION L26319
VERSION    L26319.1 GI:600756
KEYWORDS   esterase; pregastric esterase.
SOURCE     Bos taurus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1417)
AUTHORS    Timmermans,M.Y., Teuchy,H. and Kupers,L.P.
TITLE      The cDNA sequence encoding bovine pregastric esterase
JOURNAL    Gene 147 (2), 259-262 (1994)
MEDLINE    95011625
PUBMED     7926811
FEATURES   Location/Qualifiers
            1..1417
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /cell_type="Secretory cell"
            /tissue_type="Tongue serous gland"
            /dev_stage="25 weeks"
            73..1266
            /EC_number="3.1.1.3"
            /codon_start=1
            /product="pregastric esterase"

CDS
1..1272
/protein_id="AAA57037.1"
/db_xref="GI:600757"
/translation="MMWLLVTVCFTHMSGNAFCFLGKIAKNPEASMNVSQIMISYWGYP
SEMKVITADCVILQVYRIPHGKNNANHLGQRPVVFLOHGLLGSAATNWSLNPKNSLG
ELLADAGYDVLGNSRGNTWAQEHLYSPDSPEFWAFSEDEMAEYDLSESTIDFLIRRT
GOKLHYVHSGOGTTGFIATFSTPTLAERIKVFYALAPVATVATYKTSKLFNKLALIPH
FLFKIIFGDKMFYPHTLEQFLGVEMCSRETLDVLCNKALFAITGVDNKNFMSRLDY
YIAHNPAGTSVQNTLHWRAQVSKGFQADMGAPQYONLMHYHOPTPPYLYNTAMNPVI
AVWSADNLLADPDQDFLLSKLSNLYHKEIPNYNHLDFIWMADPAQEVYNEIVSLM
AEDKK"
73..129
sig_peptide
130..11263
mat_peptide
/product="pregastric esterase"
/EC_number="3.1.1.3"
/function="triacylglycerol acylhydrolase"
/evidence=experimental
1378..1385
polyA_signal
1396
polyA_site
BASE COUNT      408 a      306 c      276 g      427 t
ORIGIN

Query Match      5.9%; Score 88; DB 4; Length 1417;
Best Local Similarity 53.8%; Pred. No. 6.2e-09;
Matches 228; Conservative 0; Mismatches 190; Indels 6; Gaps 2;

QY 194 GGCTACCCCTGCACGACGACCAACGTTGAAACAAAGATGGATTCCTTTTATCTCTTCAG 253
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 196 GGCTACCCCAAGTGAGATGCATAAGTTATACTGGGATGGTTATATCTTCAGGCTCAT 255
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 254 CATATCCCAACATGGCAAAATAAAGCAGCAGAT---AGTACTGGCCCTCCAGTTTTCCT 310
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 256 CGGATTCTCATGGAAGAATAATGCTAATCATTTAGGTCAGAGACCTCTTGTGTTCTG 315
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 311 CAACATGGTCTTTTCAGGGAGGAGACACATGGTTTCATAAACTCTGCTCAGCAATCATCT 370
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 316 CAGCATGGTCTTCTTGGATCAGCCACAAACTGGATTTCCAACTTCGCCAAGAACAGCCTG 375
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 371 GGGTATATCTGCTGCATTAACGGTTTGTGATTTGGATTGGGAATGTCGCTGGAACGCGT 430
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 376 GGCTTCTCTGCAGATGCTGTTATGACGTGTGCTGGGACAGACAGAGAAACACC 435
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 431 TGGAGTAAAGGTCAATCAACCTTTTCTGTTATGATAAGCTTTTCGGAATGGAGCTGG 490
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 436 TGGGCCCAAGAACATTTATACATATTCACCAGACTCCCCGGAATTTCTGGGCTTTCAGCTTT 495
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 491 CAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGTATACAGTCAC--A 547
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 496 GATGAATGCGGGAATATGACCTTCATCTACAATTTGATTTCACTTAAGGAGAACAGGA 555
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 548 CAGTCCAAAATTTCTATATGTGGGCAATTCACAGGGAACATAATGGGTTTGGCGCTTTG 607
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 556 CAGAAGAAGCTACACTATGTGGCCATTCGCCAATCCCAAGGACACCACCATTTGTTTATCGCCTTT 615
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
QY 608 ACGA 611
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 616 TCTA 619
    ||| | | ||||| ||||| ||| | | | | | | | | | | | | | | | |

RESULT 12
BC031933
LOCUS      BC031933
DEFINITION Mus musculus, RIKEN cDNA 4632427C23 gene, clone MGC:25790
IMAGE:4020396, mRNA, complete cds.
ACCESSION BC031933
VERSION    BC031933.1 GI:21594465
KEYWORDS   MGC.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2577)
AUTHORS    Strausberg,R.
TITLE      Direct Submission
```

**JOURNAL** Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK COMMENT** NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgabs-femail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cdNA Library Preparation: Life Technologies, Inc.  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
Contact: amg@bcm.tmc.edu  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 30 Row: k Column: 3  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

**FEATURES** Location/Qualifiers  
**source**  
1..2577  
    `/organism="Mus musculus"`  
    `/db_xref="taxon:10090"`  
    `/map="CZBCH II"`  
    `/clone="MGC:25790 IMAGE:4020396"`  
    `/tissue_type="Mammary tumor metastasized to lung."`  
    MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."  
    `/clone_lib="NCI_CGAP_Lu30"`  
    `/lab_host="DH10B"`  
    `/note="Vector: pCMV-SPORT6"`  
    112..1380  
        `/codon_start=1`  
        `/product="RIKEN CDNA 4632427C23 gene"`  
        `/protein_id="AAH31933.1"`  
        `/db_xref="GI:21594466"`  
        `/db_xref="LocusId:78753"`  
        `/translation="MSEILSRVMVTSHRVEIWLLIVALLQNVNSGHLPTKAADPE AFNWSIEIKHKSGPSEEYEVATEDGYILSVNRIPRGOTRLKKESRPVVVLQHGLLG DASNTSLNPKNGPSLEIFADAGFDVWMSGRNTWSRKHTLSIDODEFAWFSDYDMA RFDLPAVINFLLOKTGOKKYVYGYSQGTIMGFIESTPELAHH1KMYPALAPIATY KYARSPGTFLPLDPMIMKVFQRBEFLTYREFRQLFYLCGGMILDQCISLIILL GGFTNNMNMSRVAVVTPACTSYONILHWSQVNSGELRAFDMSGSTKNQEKNC LPPIRYKRWMDMPVTAPGWTGGDWLSDPDVDTLLEVTNLIYKHNIPEWAHFDFWIFG LDAQPVYNIEIIHLMOEPNLPGTCRVKL"`

**CDS**  
790 a 496 c 528 g 763 t

	Query Match	Score	DB	Length
Best Local Similarity	5.9%	88;	2577;	
Matches	228; Conservative	0;	Mismatches 190; Indels 6; Gaps 2;	
Qy	194	GGCTACCCCTGCACCGACACACGGTTGAACAACAAAGATGGATTCTTTATCTCTTCAG	253	
Db	280	GTTATCCCAGTAGGAGATGATGAAGTTTGCAACAAGAAGATGGTACATCTTCTGTGTAAC	339	
Qy	254	CATATCCCATGG--CAAATAAAGCAGCACATAGTACTGCCCTCCAGTTTTCTT	310	
Db	340	AGAATCCCTCGGGGACAGACACGGTTAAAGAAAGAAAGATCCAGGCCAGTGGTGTACTG	399	
Qy	311	CAACATGGTCTTTTTCAGGAGGACACACATGGTTTATAAACTCTGCTGACCAATCACTT	370	
Db	400	CAACATGGTCTTCTGGGATGCTAGCACTGGATTTCCACACCTGCCCAACACAGCCTG	459	
Qy	371	GGGTATATCCCTGTGATAACGGTTTTTGATGTTTGGATGGGAATGTCOGTGAACCGGT	430	
Db	460	GGTTTATTTCTGCAGATGCAGGTTTGTGTTGGATGGGAAACACAGACAGAGAACACC	519	



FEATURES source Location/Qualifiers  
1. .618  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="U13071"  
/note="This clone is in pUNI 51  
ecotype: Columbia"  
1. .618  
/note="unknown protein"  
/codon\_start=1  
/product="AT5g14180/MUA22\_18"  
/protein\_id="AA010310.1"  
/db\_xref="GI:20147187"  
/translation="MLSPVAYLSHMTTVIGDIAAKTFLAEATSTILGWPEFNPKSGLVG  
DFKAICGLKIDCVLSVITGKNCLNASTIDLFLANPQSTSTKNMHLAQTVRD  
KELKYNYSGSDRNKIKHYGAIPPAYNISAIHELPLFFSYGLGSLADVKDVEFLID  
QFYHIDIKMNQFVKDIAHAFIMYITAKDVVYNQVATEFKRQA"  
BASE COUNT 175 a 144 c 130 g 169 t  
ORIGIN  
Query Match 5.8%; Score 86.6; DB 8; Length 618;  
Best Local Similarity 49.78; Pred. No. 1.1e-08;  
Matches 286; Conservative 0; Mismatches 274; Indels 15; Gaps 2;  
Qy 654 GTCCTATTCTTATCTTGATCATGTTAGTGTAGTTTGTCTTCAGAGCAGTCGCCATGC 713  
Db 8 GTCCCGTTGCTTATCTCAGCCACATGACCACCGTCATCGCGGATATCGCCGCAAAACCT 67  
Qy 714 ATCTTGATCAGATGCTGTTACTATGGAATTCACCAGCTGAACCTTCGTCAGGCATGG 773  
Db 68 TCCTCGCGCGGCCACCTCTATTCTTGATGCCAGATTTAAACCGAAAAGTGGATTAG 127  
Qy 774 GGGTTCAAATAGTAGATCTTTGTGCGATGTTGAACACGTGGATTGCAACAATTTGCTAT 833  
Db 128 TAGGGATTTTATAAGCTATATGCTTTAAAGCGGGATCGACTGTTATGATTAGTCT 187  
Qy 834 CTGGGATTACAGGGAAACATGTTGCTTCAATACATCAAGGATTGATTATTATTGGAGT 893  
Db 188 CTGTTATCACTGGAAAGAAATGCTGCCTTAACGCATCAACCATTTGCTCTCTTGCCA 247  
Qy 894 ATGAACCTCATCCATCATCGACAAAAATCTGCACCATCTTTTCAGATGATCAGGAAG 953  
Db 248 ACGAACCTCAATCTACTTCCACCAAGAACATGATGCACCTTGCTCAACCGTAAGACAC 307  
Qy 954 GCACCTTCCGAAAGATGACTATGGTTATTGG---GAAACCTAAGCGCTACGTCATT 1010  
Db 308 AGGAGTTGAGAAATACAACTATGGAAGCAGTGTGCGCAACATTAACATTACGCTCAAG 367  
Qy 1011 TCGCTCTCCCGCATTTGACCTAAGCAGCATACCAGAACTACTGCCCATATGGATGGAT 1070  
Db 368 CGATACCAACCGCATACAAATATCGGGGATCCACACGAGCTCCGCTTTTTCAGCT 427  
Qy 1071 ATGGAGTCTTGATGCGATGCTGATGTAACCGATGTTCCAGCGTACTATCAGAGCTGG 1130  
Db 428 ACGGTGGTTAGATAGCTGCTGATGTAAGAGCGTTGAGTTCTCTCTTGACCAGTTTA 487  
Qy 1131 GAT-----CTACACCAAGACTTCTGTACATTTGTTGCTATGCGCATATTTGATT 1178  
Db 488 AGTATCAGCAGATCGATGAAGATGAATGTGCGAGTTTGTGAAGACTATGCTCATGCTGATT 547  
Qy 1179 TTGTTATGAGCGTGAAGCGCAAGATGATGTTTAT 1213  
Db 548 TCATCATGGGTGTTACTGCCAAAGACGTTGTTTAT 582

RESULT 14  
AX392821  
LOCUS AX392821  
DEFINITION Sequence 7 from Patent W00216597.  
ACCESSION AX392821  
VERSION AX392821.1 GI:19700917  
KEYWORDS

AX392821  
DEFINITION Sequence 7 from Patent W00216597.  
ACCESSION AX392821  
VERSION AX392821.1 GI:19700917  
KEYWORDS

AX392821  
DEFINITION Sequence 7 from Patent W00216597.  
ACCESSION AX392821  
VERSION AX392821.1 GI:19700917  
KEYWORDS

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Griffin,J.A., Patterson,C., Gandhi,A.R., Lu,Y., Yao,M.G.,  
Baughn,M.R., Walla,N.K., Hafalia,A.J., Ding,L., Tribouley,C.M.,  
Das,D., Thornton,M. and Lal,P.  
TITLE Lipid metabolism enzymes  
JOURNAL Patent: WO 0216597-A 7 28-FEB-2002;  
FEATURES Location/Qualifiers  
source 1. .1228  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7479063CB1"  
BASE COUNT 376 a 244 c 259 g 349 t  
ORIGIN  
Query Match 5.8%; Score 86.4; DB 6; Length 1228;  
Best Local Similarity 53.58; Pred. No. 1.4e-08;  
Matches 227; Conservative 0; Mismatches 191; Indels 6; Gaps 2;  
Qy 194 GGCTACCCCTGACCGAGCACAACGTTGAAACAAAGATGATTCCTTTTATCTCTTCAG 253  
Db 145 GGCTACCCCGAGTGAAGAGTATGAAGTCACTGAGATGGTATATCTCTCTGCAAC 204  
Qy 254 CATATCCACATGGCAAAATAAAGCAGCAGATAGTACTGGCCCTCCAGTT---TTTTCTT 310  
Db 205 AGAATTCCTTATGGCGAACACATGCTAGGAGCAGAGTCCCGCCAGTTGTGTATATG 264  
Qy 311 CAACATGCTCTTTTTCAGGGAGGAGACACATGTTTCATAAACTCTGCTGAGCAATCACTT 370  
Db 265 CAGCATGCCCTGTTTGCAGGACAGCCCTACTGGCTTGAGAATTATGCCAATGGAAGCCTT 324  
Qy 371 GGGTATATCTTGTGATPAACGGTTTTCATGTTGGATTGGGAATGTCGGTGGAAACGGT 430  
Db 325 GGATTCTCTTACGAGATGACAGTTATGATGATGGTGGAAACAGTCGGGGAACACT 384  
Qy 431 TGGACTAAGGTCATTCACACCTTTTCTGTTTCATGATAGCTTTTCTGGGATTTGGAGTGG 490  
Db 395 TGGTCAAGAAGACACAAACACTCTCAGAGACAGATGAGAAATCTCGGCCCTTTAGTTT 444  
Qy 491 CAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGCT---ATGCTGTATACAGTCACA 547  
Db 445 GATGAATGGCCAAATATGATCTCCAGAGTAATAGACTTCATTGTAATAAAACTGCT 504  
Qy 548 CAGTCCAAAATTCATATGTTGGGCAATTCACAGGAACTATAATGGGTTTGGCGGCTTTG 607  
Db 505 CAGGAGAAATTTGATTTTCATTGGACATTCCTTGGCACTACAAATAGGCTTTGTAGCCTT 564  
Qy 608 ACGA 611  
Db 565 TCCA 568  
RESULT 15  
CBRG44C02/C  
LOCUS CBRG44C02 41805 bp DNA linear INV 04-NOV-2000  
DEFINITION Caenorhabditis briggsae cosmid G44C02, complete sequence.  
ACCESSION AC084610  
VERSION AC084610.1 GI:11095059  
KEYWORDS HTG.  
SOURCE Caenorhabditis briggsae.  
ORGANISM Caenorhabditis briggsae.  
REFERENCE 1  
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
TITLE 1 (bases 1 to 41805)  
JOURNAL Washington University Genome Sequencing Center.  
REFERENCE 2 (bases 1 to 41805)  
UNPUBLISHED The C. briggsae Genome Sequencing Project  
AUTHORS Wilson,R.

TITLE The sequence of C. briggsae cosmid G44C02  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 41805)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (04-Nov-2000) Department of Genetics, Washington  
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: jspliehwatson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

FEATURES  
source  
1. .41805  
Location/Qualifiers  
/organism="Caenorhabditis briggsae"  
/strain="Gujarat G16"  
/db\_xref="taxon:6238"  
/clone="G44C02"

BASE COUNT 13098 a 8049 c 8106 g 12552 t  
ORIGIN

Query Match 5.8%; Score 85.8; DB 3; Length 41805;  
Best Local Similarity 54.7%; Pred. No. 3.6e-08;  
Matches 193; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
Qy 242 TTAATCTCTTACGACATCCACATGGCAAAATAAAGCAGCAGATAGTACTGGCCCTCCA 301  
Db 29801 TTGTCCTATGCCCTAATACCATATTCAGAGAAATGCCATACAAAGTCAAAACGTCCAATT 29742  
Qy 302 GTTTTCTTCAACATGGTCTTTTTCAGGGAGGAGACACATGGTTTCATAAAACTCTGCTGAG 361  
Db 29741 GTATTTTACAAACATGGTCTGCTAGCAGATGGATTTTCATGGATTCCAAATTTGGCTAAT 29682  
Qy 362 CAATCACTTGGGTATATCCTTGCTGATAACGGTTTTCATGTTGGATTGGGAATGTCGGT 421  
Db 29681 CAATCAGCTGGATTGTTTTCGCGATGCAGGGTTTTCATGTTGGATTGCCAATAGTCGG 29622  
Qy 422 GGAACGGCTTGGAGTAAAGGTCATTCAACCTTTCTGTTATGATAAGCTTTTCTGGGAT 481  
Db 29621 GGTACTCCAGCATCACAAAGCATATCGGATATGCTCCAGAAATCAGAAATCTGGAAT 29562  
Qy 482 TGGAGCTGGCAAGAGTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGCTATACA 541  
Db 29561 TTCATATGGCAACAAATGCTGTAATAGCATTTGACTTCGAGTGTGATTACGCTCTGAAA 29502  
Qy 542 GTCACACAGTCCAAATCTA--TATCTGGGGCATTACAGGGGAACATAAT 591  
Db 29501 GAGACAAACAAAGTTTTTATACTATTGGGACATTCTCAAGGAACATATGAT 29449

Search completed: December 30, 2002, 18:29:11  
Job time : 3950 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2002, 17:10:06 ; Search time 281 Seconds  
(without alignments)  
11885.101 Million cell updates/sec

Title: US-09-699-652A-13  
Perfect score: 1483  
Sequence: 1 gcacgagtacacgagcggc.....ctctcaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Maximum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1483	100.0	1483	21	Rice acid triacylg
2	529.4	35.7	859	21	Corn acid triacylg
3	445.6	30.0	1438	21	Soybean acid triac
4	184.6	12.4	1718	21	Soybean acid triac
5	94.2	6.4	802	22	Human neuroblastom
6	94.2	6.4	2481	22	Gene #3881 used to
7	94.2	6.4	2493	22	Human shear stress
8	91.2	6.1	1411	24	Kid goat pregastric
9	91.2	6.1	8324	24	Yeast YE-1 express

10	89.4	6.0	1116	24	ABK86570
11	89.4	6.0	1210	24	ABK86569
12	89.4	6.0	1269	22	AAF45132
13	89.4	6.0	1384	22	AAAD19220
14	89.4	6.0	2035	22	AAAD19226
15	89.4	6.0	2044	22	AAF45131
16	86.4	5.8	1228	24	ABK12385
17	83.8	5.7	417	22	AAH50931
18	83.8	5.7	417	22	AAH56938
19	83.2	5.6	1284	24	ABK85978
20	81.6	5.5	1360	24	AAAL7480
21	81.6	5.5	1695	22	AAH57420
22	80	5.4	617	21	AAA44349
23	80	5.4	1367	7	AAH60566
24	80	5.4	1367	7	AAH60566
25	80	5.4	1367	17	AAH58916
26	79.6	5.4	705	21	AAZ34952
27	78.4	5.3	1378	14	AAQ42310
28	77.6	5.2	1211	23	ABL29771
29	77.6	5.2	3442	23	ABL29770
30	76.8	5.2	1336	6	AAH50385
31	71	4.8	1242	23	ABL06473
32	71	4.8	4196	23	ABL05472
33	68.6	4.6	1473	24	ABK12388
34	67.2	4.5	1185	24	AAH31193
35	66.2	4.5	1197	23	ABL12389
36	66.2	4.5	3269	23	ABL12388
37	65.8	4.4	1098	24	AAH27802
38	65.4	4.4	1200	22	AAH78205
39	65.2	4.4	735	22	AAAL2482
40	64.4	4.3	1528	15	AAQ68388
41	64.4	4.3	1528	17	AAH58915
42	64.4	4.3	1531	15	AAQ68389
43	62.4	4.2	1194	24	AAH27803
44	60.8	4.1	699	22	AAH28679
45	60.8	4.1	828	24	AAH31194

ALIGNMENTS

RESULT 1  
AAZ34956  
ID AAZ34956 standard; cDNA; 1483 BP.  
XX AC AAZ34956;  
XX AC  
DT 28-FEB-2000 (first entry)  
XX DE Rice acid triacylglycerol lipase cDNA.  
XX KW Triacylglycerol lipase; rice; fatty acid; seed oil;  
XX KW vegetable oil; transgenic plant; ss.  
XX OS Oryza sativa.  
XX FH Key  
XX CDS Location/Qualifiers  
FT 32..1264  
FT /\*tag= a  
XX PN WO9955883-A2.  
XX PD 04-NOV-1999.  
XX PF 29-APR-1999; 99WO-US09280.  
XX PR 30-APR-1998; 98US-0083688.  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX PI Cahoon EB, Cahoon RE, Kinney AJ, Rafalski JA;  
XX DR WPI; 2000-062036/05.

DR P-PSDB; AAY32307.  
XX  
PT Novel plant triacylglycerol lipase polynucleotides used to alter the  
XX level of the enzyme in transgenic plants -  
PS Claim 2; Page 44; 65pp; English.  
XX  
CC This is the nucleotide sequence of the cDNA insert in clone  
CC rlr72.pk0015.b2 encoding the entire rice acid triacylglycerol  
CC lipase (TAGL) (see AAY32307). The clone was isolated from a rice  
CC leaf cDNA library. Novel acid and neutral TAGL polypeptides  
CC (see AAY32301-17) and polynucleotides (see AAZ34950-66) from corn,  
CC Catalpa, rice, soybean and wheat tissues are disclosed. The enzymes  
CC may be prepared recombinantly and used to raise antibodies, which  
CC are used for detecting the enzymes in situ in cells or in vitro in  
CC cell extracts. The polynucleotides may be used to create transgenic  
CC plants in which the TAGL levels are present at higher or lower levels  
CC than normal, or in cell types or developmental processes where they are  
CC not normally found. This would alter the level of triacylglycerol and  
CC cholesterol esters found in those cells. Accumulation of fatty acids  
CC with unusual structures may be a positive phenotype in plants used for  
CC foods. In addition, it may be desirable to eliminate expression of TAGL  
CC genes for certain applications. TAGL enzymes may also be useful for the  
CC processing of plant seed oils and for the development of novel seed  
CC oils. The TAGL enzymes can also be used as targets to facilitate the  
CC design and/or identification of inhibitors of those enzymes that may be  
CC useful as herbicides. This is desirable because inhibition of the  
CC activity of either of the enzymes could lead to an inhibition of plant  
CC growth. The polynucleotides also serve as a source of probes and  
CC primers, which are useful for genetic mapping, as markers for traits  
CC linked to those genes, and to isolate homologous sequences from other  
CC species.  
XX  
SQ Sequence 1483 BP; 375 A; 334 C; 344 G; 430 T; 0 other:  
Query Match 100.0%; Score 1483; DB 21; Length 1483;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCACGAGTACACAGCGCGCGGCGTGGCGATGGCGATGGCGGCGCACGCCCCCGGAGG 60  
DB 1 GCACGAGTACACAGCGCGCGGCGTGGCGATGGCGATGGCGGCGCACGCCCCCGGAGG 60  
QY 61 AGCGTCCCGCTGATCTCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 61 AGCGTCCCGCTGATCTCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 121 CCCAGCGCGCGCGCGCTCCGCGCGTCCGCGTCCGCGCGCTCCGCGCGCTCCGCGCGCT 180  
DB 121 CCCAGCGCGCGCGCGCTCCGCGCGTCCGCGTCCGCGCGCTCCGCGCGCTCCGCGCGCT 180  
QY 181 GCTCTGCGACATCGGCTACCGCTGACCGAGCACACAGCTTGAACAAAGATGGATTCTT 240  
DB 181 GCTCTGCGACATCGGCTACCGCTGACCGAGCACACAGCTTGAACAAAGATGGATTCTT 240  
QY 241 TTATCTCTTCAGCATATCCACATGGCAAAATAAGCAGCAGATAGTACTGGCCCTCC 300  
DB 241 TTATCTCTTCAGCATATCCACATGGCAAAATAAGCAGCAGATAGTACTGGCCCTCC 300  
QY 301 AGTTTTTCTTCAACATGTTCTTTTTCAGGAGGAGACACATGGTTTCATAAACTCTGCTGA 360  
DB 301 AGTTTTTCTTCAACATGTTCTTTTTCAGGAGGAGACACATGGTTTCATAAACTCTGCTGA 360  
QY 361 GCAATCATTGGGTATATCTCTGCTGATACCGTTTTTGATGTTGGATGGGAATGTCGG 420  
DB 361 GCAATCATTGGGTATATCTCTGCTGATACCGTTTTTGATGTTGGATGGGAATGTCGG 420  
QY 421 TGAACGCGCTTGAGTAAAGGTCATTCAACCTTTTCTGTTTCATGATAGCTTTTCTGGGA 480  
DB 421 TGAACGCGCTTGAGTAAAGGTCATTCAACCTTTTCTGTTTCATGATAGCTTTTCTGGGA 480  
QY 481 TTGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGTATATAC 540  
DB 481 TTGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGTATATAC 540

DB 481 TTGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGTGTATATAC 540  
QY 541 AGTCACACAGTCCAAAATTCATATATGTGGGCAATTCACAGGGAACATATAATGGGTTTGGC 600  
DB 541 AGTCACACAGTCCAAAATTCATATATGTGGGCAATTCACAGGGAACATATAATGGGTTTGGC 600  
QY 601 GCGTTTGGAGATGCCGAAATAGTAAATAGTATAGCTCTGCAGCACCTTCTTTGTGCTCTAT 660  
DB 601 GCGTTTGGAGATGCCGAAATAGTAAATAGTATAGCTCTGCAGCACCTTCTTTGTGCTCTAT 660  
QY 661 TTCTTATCTTGATCATGTTAGTGTAGTTTGTCTCAGAGCAGTCGCCATCATCTTTGA 720  
DB 661 TTCTTATCTTGATCATGTTAGTGTAGTTTGTCTCAGAGCAGTCGCCATCATCTTTGA 720  
QY 721 TCAGATGCTTGTACTATGGGAATTCACAGCTGAACCTTCCTAGCAGCATGGGGTTCA 780  
DB 721 TCAGATGCTTGTACTATGGGAATTCACAGCTGAACCTTCCTAGCAGCATGGGGTTCA 780  
QY 781 AATAGTAGATCTTTTGTGCGATGGTGAACACGTGATTCGAACAATTTGCTATCTGCGAT 840  
DB 781 AATAGTAGATCTTTTGTGCGATGGTGAACACGTGATTCGAACAATTTGCTATCTGCGAT 840  
QY 841 TACAGGGGAAAACCTGCTTCAATACATCAAGGATTCATTTATTTGGAGTATGAACC 900  
DB 841 TACAGGGGAAAACCTGCTTCAATACATCAAGGATTCATTTATTTGGAGTATGAACC 900  
QY 901 TCATCATCATCGACAAAATCTGCACCATCTTTTCAGATGATCAGGAAGGACATTT 960  
DB 901 TCATCATCATCGACAAAATCTGCACCATCTTTTCAGATGATCAGGAAGGACATTT 960  
QY 961 CGCAAGATGATGATGGTTTATGGGAAACCTTAGCGGCTACGGTCAITTCGCTCTCC 1020  
DB 961 CGCAAGATGATGATGGTTTATGGGAAACCTTAGCGGCTACGGTCAITTCGCTCTCC 1020  
QY 1021 CGCATTTGACCTTAAGCAGCATACCAATCACTGCCCATATGATGGGATATGGAGTCT 1080  
DB 1021 CGCATTTGACCTTAAGCAGCATACCAATCACTGCCCATATGATGGGATATGGAGTCT 1080  
QY 1081 TGATGATGCTGATGATTAACCGATGTTTACGGGTACTATCAGAGAGCTGGGATCTACACC 1140  
DB 1081 TGATGATGCTGATGATTAACCGATGTTTACGGGTACTATCAGAGAGCTGGGATCTACACC 1140  
QY 1141 AGAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 1141 AGAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
QY 1201 AGATGATGTTTATGAGACCTAATAGATTTCTTAGGGAATAGGATGGCATTAATAGCTA 1260  
DB 1201 AGATGATGTTTATGAGACCTAATAGATTTCTTAGGGAATAGGATGGCATTAATAGCTA 1260  
QY 1261 TTAGGATGCTTCATGCTGATATAATAAACAATCTGTACAGTATTTGCTCTCCGATGT 1320  
DB 1261 TTAGGATGCTTCATGCTGATATAATAAACAATCTGTACAGTATTTGCTCTCCGATGT 1320  
QY 1321 GAGTATGATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
DB 1321 GAGTATGATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
QY 1381 AGCTGTGACGAGCAATTTGATCATTTGATCCAACTTATCGCTCCCACTACTGTATATCCA 1440  
DB 1381 AGCTGTGACGAGCAATTTGATCATTTGATCCAACTTATCGCTCCCACTACTGTATATCCA 1440  
QY 1441 TTATAGAAACCCCTCTTCATTTCTCTCAAAAAAAAAAAAAA 1483  
DB 1441 TTATAGAAACCCCTCTTCATTTCTCTCAAAAAAAAAAAAAA 1483

RESULT 2  
AAZ34953  
ID AAZ34953 standard; cDNA: 859 BP.  
XX  
XX AAZ34953;  
XX













```
Query Match      6.1%; Score 91.2; DB 24; Length 1411;
Best Local Similarity 54.2%; Pred. No. 2.3e-14;
Matches 230; Conservative 0; Mismatches 188; Indels 6; Gaps 2;

QY 194 GGTACCCCTGCACGAGCACAACGTTGAAACAAGATGATCCCTTTATCTCTTCAG 253
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 191 GGCTACCCAAGTGAGATGCATAAAGTTATACTGCAGATGGCTATATCTTCAGGTCTAT 250
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 254 CATATCCACATGCGCAAAATAAAGCAGCAGAT---AGTACTGGCCCTCCAGTTTCTT 310
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 251 CGGATTCCTCATGGAAAGATGATGCTAATCATTTAGTTCAGAGACCTGTGTGTTCTG 310
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 CAACATGCTCTTTTTCAGGAGGAGACACATGTTTCATATAAATCTGTCGACCAATCACT 370
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 311 CAGCATGGTCTTCTTGGCTCAGCTACAACTGGAATTTCCAACTTCCCAACAACAGCCTG 370
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 371 GGGTATATCTTGCTGATAACGTTTTGATGTTGGATTTGGAAATGTCCTGGGAACGCT 430
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 371 GGCCTTCCCTCGGAGATGCTGGTTATGACGTGTGGCTGGGGAACAGCAGAGAAACACT 430
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 431 TGGAGTAAAGTTCATTAACCTTTTCTGTTATGATAAGCTTTTCTGGGATTGGAGCTGG 490
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 431 TGGGCCCAGGACATTTTACTATTCACCACTCCCTGAACTTCTGGGCTTCAGCTTT 490
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 491 CAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGCTATGCT---ATACAGTCACA 547
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 491 GATGAATGGCTGAATATGACCTTCCATCTACAATTGATTTTCATCTTAAAGAAACAGGA 550
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 548 CAGTCCAAATTTCTATATGTGGGGCAATTCACAGGGAATATAATGGGTTTGGCGCTTTG 607
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 551 CAGAAGAAGCTACACTATGTGGCCATTCCTCCCAAGCACCACCATTTGGTTTGTGCGCTTT 610
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 608 ACGA 611
DB 611 TCTA 614

RESULT 9
ABL57145
ID ABL57145 standard; DNA; 8324 BP.
XX
AC ABL57145;
XX
DT 05-AUG-2002 (first entry)
XX
DE Yeast YE-1 expression vector.
XX
KW Vector; YE-1; esterase; lipase; goat; cheese; flavour; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN US6375947-B1.
XX
PD 23-APR-2002.
XX
PF 05-NOV-1998; 98US-0186489.
XX
PR 05-NOV-1998; 98US-0186489.
XX
PA (INFL ) INT FLAVORS & FRAGRANCES INC.
XX
PI Bolen PL, Cihak PL, Scharpf LG;
XX
DR WPI; 2002-424751/45.
XX
PT An isolated kid goat pregastric esterase free from other goat proteins
PT is useful to prepare enzyme modified cheese and to treat animals with
PT an esterase or lipase deficiency
XX
PS Disclosure; Column 29-38; 35pp; English.
XX
CC The present sequence is that of yeast YE-1 expression vector.
XX
CC The vector comprising a nucleic acid (see ABL57144) encoding kid
```

```
CC goat pregastric esterase (kPGE, see ABB76189) can be used to
CC transform a host cell, such as Saccharomyces cerevisiae, for
CC recombinant production of kPGE. The recombinant kPGE can be used
CC to treat animals with an esterase or lipase deficiency. The enzyme
CC produces a mixture of fatty acids when reacted with a dairy product
CC such as lipolysed butter oil, milk, cheese or whey. It is
CC therefore useful in the production of enzyme modified cheeses as a
CC substitute for rennet, and in the creation of flavour in cheese
CC foods or any other food containing triglycerides. Microbial
CC production will allow the development of new kosher and vegetarian
CC food products.
XX
SQ Sequence 8324 BP; 2394 A; 1884 C; 1703 G; 2343 T; 0 other;

Query Match      6.1%; Score 91.2; DB 24; Length 8324;
Best Local Similarity 54.2%; Pred. No. 5.7e-14;
Matches 230; Conservative 0; Mismatches 188; Indels 6; Gaps 2;

QY 194 GGTACCCCTGCACGAGCACAACGTTGAAACAAGATGATCCCTTTATCTCTTCAG 253
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1598 GGCTACCCAAGTGAGATGCATAAAGTTATACTGCAGATGGCTATATCTTCAGGTCTAT 1657
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 254 CATATCCACATGCGCAAAATAAAGCAGCAGAT---AGTACTGGCCCTCCAGTTTCTT 310
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1658 CGGATTCCTCATGGAAAGATGATGCTAATCATTTAGTTCAGAGACCTGTGTGTTCTG 1717
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 CAACATGCTCTTTTTCAGGAGGAGACACATGTTTCATAAATCTGCTGAGCAATCACTT 370
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1718 CAGCATGGTCTTCTTGGCTCAGCTACAACTGGATTTCCAACTTCCCAACAACAGCCTG 1777
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 371 GGGTATATCTTGCTGATAACGTTTTTGTATGTTGGGAATGTCCTGGGAACGCTG 430
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1778 GGTCTCTCTGGCAGATGCTGGTTATGACGTGTGGCTGGGAACAGCAGGAACACT 1837
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 431 TGGAGTAAAGTCAATTAACCTTTTCTGTCATGATAAGCTTTTCTGGGATTGGAGCTGG 490
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1838 TGGGCCCAAGAACATTTATCTATTCACCACTCCCTGAACTTCTGGGCTTTCAGCTTT 1897
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 491 CAAGAGTTAGCTGAATATGACCTTTTACCAATGCTAGCTATGCT---ATACAGTCACA 547
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1898 GATGAATGGCTGAATATGACCTTCCATCTACAATTGATTTTCATCTTAAAGAAACAGGA 1957
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 548 CAGTCCAAATTTCTATATGTGGGCAATTCACAGGGAATATAATGGGTTTGGCGCTTTG 607
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1958 CAGAAGAAGCTACACTATGTGGCCATTCCTCCCAAGCACCACCATTTGGTTTGTGCGCTTT 2017
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 608 ACGA 611
DB 2018 TCTA 2021

RESULT 10
ABK86570
ID ABR86570 standard; cDNA; 1116 BP.
XX
AC ABR86570;
XX
DT 24-SEP-2002 (first entry)
XX
DE Human cDNA encoding lysosomal acid lipase #2.
XX
KW Human; ss; gene; lysosomal acid lipase; lipid malabsorption illness;
KW cystic fibrosis; alcoholism; heart disease; heart attack;
KW Wolman disease; cholesterol ester storage disease; brain injury;
KW mood disorder; anxiety disorder; thought disorder; volition disorder;
KW sleep disorder; neurogenic disorder; myopathic disorder; COPD;
KW obesity; cancer; chronic obstructive pulmonary disease; diabetes;
KW cardiovascular disorder; Alzheimer's disease; Parkinson's disease;
KW anorexia; osteoarthritis; central nervous system disorder;
KW peripheral nervous system disorder.
XX
OS Homo sapiens.
XX
```



CC treatment of a disease such as lipid malabsorption illness,  
CC cystic fibrosis, alcoholism, heart disease, heart attack,  
CC Wolman disease, cholesterol ester storage disease, brain injury,  
CC mood disorder, anxiety disorder, thought disorder, volition disorder,  
CC sleep disorder, neurogenic disorder, myopathic disorder, obesity, cancer,  
CC chronic obstructive pulmonary disease (COPD), diabetes,  
CC cardiovascular disorder, Alzheimer's disease, Parkinson's disease,  
CC anorexia, osteoarthritis, a central nervous system disorder and  
CC a peripheral nervous system disorder. The present sequence is the  
CC cDNA encoding human lysosomal lipase #1.  
XX  
SQ Sequence 1210 BP; 338 A; 264 C; 296 G; 312 T; 0 other;  
Query Match 6.0%; Score 89.4; DB 24; Length 1210;  
Best Local Similarity 53.6%; Pred. No. 6.5e-14;  
Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;  
QY 183 TCCTGCCACTGGCTACCCCTGCACCGAGCACACAGTGAACAAAGATGGATTCCTTT 242  
DB 83 TCCAAACATCAAGGCTATCCCTGTGAGGAATATGAAGTCGCAACTCAAGATGGGTATATCC 142  
QY 243 TATCTCTTCAGCATATCCACATGTCGCAAAATAAGCAGCAGATA--GTACTGCCCTC 299  
DB 143 TTTCTGTTAAACAGGATCTCGAGCCCTAGTGCACCTTAAGAAGACAGGTTCCAGGCCGTG 202  
QY 300 CAGTTTCTTCAACATGGTCTTTTTCAGGAGGAGACACATGTTTCAATAACTCTGCTG 359  
DB 203 TGGTCTTACTGCAGCATGGCTAGTTGGAGGTGCTAGCACTGGATTTCCAACTGCCCA 262  
QY 360 ACCAATCACTTGGGTATATCTTCTGCTGATACGGTTTGTGATGTTGGATGGGAATGCC 419  
DB 263 ACAATAGGCTGGCTTCATCTGCGAGATGCTGTTTTCAGCTGTGGATGGGAACAGCA 322  
QY 420 GTGGAAACGCTTGGAGTAAAGTCTAATCAACCTTTCTGTTTCATGATAAGCTTTCTGGG 479  
DB 323 GGGGAACGCTGCTCGAAACACAGACACACTCTCCATAGACCACAGATCAGTCTCGG 382  
QY 480 ATGGAGCTGGCAAGAGTTAGTGAATATGACCTTTTACCAATGCTAGGCT---ATGTTG 536  
DB 383 CTTTCAGTTATGATGAGATGGCTAGGTTTTCAGCTTCTCTGCAAGTGAATAAATTTATTTCG 442  
QY 537 ATACAGTCACAGTCCAAATCTATATGTCGGGCATTCACAGGAACTATATGGGTT 596  
DB 443 AGAAACGGGCGGCAAGAAAGATCTATTATGTCGGCTATTTCACAGGGCCACCACCATGGGCT 502  
QY 597 TGGCGGCTTTGACGA 611  
DB 503 TTATTGCATTTTCCA 517  
RESULT 12  
AAF45132  
ID AAF45132 standard; cDNA; 1269 BP.  
XX  
AC AAF45132;  
XX  
DT 30-MAR-2001 (first entry)  
XX  
DE Human TANGO 294 ORF.  
XX  
KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;  
KW central nervous system; focal brain disorder; bipolar affective disorder;  
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;  
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;  
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;  
KW neuropsychiatric; psychoactive substance use; anxiety; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200077239-A2.  
XX  
PD 21-DEC-2000.  
XX

PF 24-MAY-2000; 2000WO-US14858.  
XX  
PR 14-JUN-1999; 99US-0333159.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;  
XX WPI: 2001-032313/04.  
DR P-PSDB; AAB66065.  
XX  
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
PT screening assays and diagnostic assays and for the treatment of  
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
PT disease.  
PS  
PS Claim 1; Fig 6; 359pp; English.  
XX  
CC The present invention relates to TANGO or INTERCEPT proteins and coding  
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,  
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
CC sequences are useful for the treatment of neurological disorders such as  
CC central nervous system (CNS) disorders, CNS-related disorders, focal  
CC brain disorders, global-diffuse cerebral disorders and other  
CC neurological and cerebrovascular disorders. The CNS disorders include  
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
CC autonomic function disorders such as hypertension and sleep disorders,  
CC neuropsychiatric disorders, psychoactive substance use disorders,  
CC anxiety, and bipolar affective disorder.  
XX  
SQ Sequence 1269 BP; 358 A; 268 C; 309 G; 334 T; 0 other;  
Query Match 6.0%; Score 89.4; DB 22; Length 1269;  
Best Local Similarity 53.6%; Pred. No. 6.6e-14;  
Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;  
QY 183 TCCTGCCACTGGCTACCCCTGCACCGAGCACACAGTGAACAAAGATGGATTCCTTT 242  
DB 158 TCCAAACATCAAGGCTATCCCTGTGAGGAATATGAAGTCGCAACTCAAGATGGGTATATCC 217  
QY 243 TATCTCTTCAGCATATCCACATGTCGCAAAATAAGCAGCAGATA--GTACTGCCCTC 299  
DB 218 TTTCTGTTAAACAGGATTCCTCGAGGCCCTAGTGCACCTTAAGAAGACAGGTTCCAGGCCGTG 277  
QY 300 CAGTTTCTTCAACATGGTCTTTTTCAGGAGGAGACACATGTTTCAATAACTCTGCTG 359  
DB 278 TGGTGTACTGCAGCATGGCCCTAGTTGGAGGTGCTAGCACTGGATTTCCAACTGCCCA 337  
QY 360 AGCAATCACTTGGGTATATCCTTCTGATAACGGTTTGTATGTTGGATGGGAATGTCC 419  
DB 338 ACAATAGCTGGGCTTCATCTCGCAGATGCTGGTTTTCAGCTGTGGATGGGAACAGCA 397  
QY 420 GTGGAACGCTTGGAGTAAAGTCAATCAACCTTTCTGTTTCATGATAGCTTTTCTGGG 479  
DB 398 GGGGAACGCTTGGTCTCGAAACACAAAGACTCTCCATAGACCAAGATGAGTCTCGG 457  
QY 480 ATGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCT---ATGTTG 536  
DB 458 CTTTCAGTTATGATGAGATGGCTAGGTTTTCAGCTTCTCGAGTGAATAACTTTATTTCG 517  
QY 537 ATACAGTCACAGTCCAAATCTTATATGTTGGGCATTCACAGGAACTATATGGGTT 596  
DB 518 AGAAACGGGCGGCAAGAAAGATCTATTATGTCGGCTATTTCACAGGGCCACCACCATGGGCT 577  
QY 597 TGGCGGCTTTGACGA 611  
DB 578 TTATTGCATTTTCCA 592  
RESULT 13  
AAD19220  
ID AAD19220 standard; DNA; 1384 BP.

XX AC AAD19220;  
 XX 18-DEC-2001 (first entry)  
 XX DE Human CG162 (or C59) lipase DNA #1.  
 XX KW Human; apolipoprotein; lipase; lipoprotein receptor; ALLr; angina;  
 KW cardiovascular disease; lipid metabolism; myocardial infarction;  
 KW cerebral ischaemia; arterial thrombosis; thrombolytic; antilipaeamic;  
 KW coronary artery thrombosis; cerebral artery thrombosis; stroke;  
 KW intracardiac thrombosis; gene therapy; cardiovascular; vasodilator;  
 KW neuroprotectant; cerebroprotective; ds.  
 XX OS Homo sapiens.  
 XX PN WO200179446-A2.  
 XX PD 25-OCT-2001.  
 XX PF 16-APR-2001; 2001WO-US12529.  
 XX PR 14-APR-2000; 2000US-197137P.  
 XX PR 20-JUN-2000; 2000US-0598042.  
 XX PR 03-AUG-2000; 2000US-0631451.  
 XX PR 22-SEP-2000; 2000US-0667298.  
 XX PR 17-NOV-2000; 2000US-0714936.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Ballinger DG, Loeb D, Montgomery JR, Tang TY, Zhou P, Goodrich R;  
 PI Liu C, Asundi V, Zhao QA, Wehrman T, Drmanac RT, Ren F, Qian XB;  
 PI Wang D;  
 XX DR WPI; 2001-611724/70.  
 XX PT Nucleic acids encoding human apolipoproteins, lipases, and lipoprotein  
 PT receptor polypeptides, useful for preventing diagnosing and treating  
 PT lipid metabolism disorders, thrombosis and cardiovascular diseases -  
 XX Claim 1; Page 160-161; 266pp; English.  
 XX CC The invention relates to polynucleotides encoding proteins CG122, CG179,  
 CC CG95, CG121, CG162, CG27, CG153 and CG168 which are related to proteins  
 CC involved in lipid metabolism and cardiovascular disease such as human  
 CC apolipoproteins, lipases and lipoprotein receptor proteins. These DNA  
 CC and protein sequences are useful for treating or preventing disorders  
 CC associated with apolipoproteins, lipases and lipoprotein receptor (ALLr)  
 CC expression and for treating lipid metabolism, cardiovascular diseases  
 CC and thrombosis. Antibodies against these proteins are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of these sequences. ALLr polypeptides are also  
 CC useful for identifying agents (agonists and antagonists) that bind to  
 CC them and cells expressing ALLr proteins are useful for identifying a  
 CC therapeutic agent for use in treatment of a pathology related to  
 CC aberrant expression or physiological interactions of this polypeptide.  
 CC Vectors comprising these DNA and protein sequences are also useful for  
 CC producing ALLr proteins. The nucleic acids and polypeptides of the  
 CC invention are also useful for the treatment of occlusive cardiovascular  
 CC diseases, myocardial infarction, cerebral ischaemia, angina, arterial  
 CC thrombosis, coronary artery thrombosis and cerebral artery thrombosis  
 CC or intracardiac thrombosis and stroke. The nucleotides of the invention  
 CC are used in gene therapy. The present sequence is human CG162 (or C59)  
 CC lipase DNA.  
 XX SQ Sequence 1384 BP; 404 A; 281 C; 328 G; 371 T; 0 other;

Query Match 6.0%; Score 89.4; DB 22; Length 1384;  
 Best Local Similarity 53.6%; Pred. No. 6.9e-14;  
 Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;

QY 183 TCCTGCCATCGGCTACCCCTCCACCGACACACGTTGAAACAAAGATGATTCCTTT 242  
 ||| | | ||||| ||||| || | | | | | ||||| | |

Db 304 TCCAACATCAAGGCTATCCCTGTGAGGAATATGAAGTCGCAACTGAAGATGGGTATATCC 363  
 QY 243 TATCTCTTCAGCATATCCACATGGCAAAATAAAGCAGCAGATA--GTAAGTGGCCCTC 299  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 364 TTTCTGTTAACAGGATTCCTCGAGGCTAGTGCACCTAAGAAGACAGGTTCCAGGCCCTG 423  
 QY 300 CAGTTTCTTCAACATGCTCTTTTTCAGGAGGAGACACATGGTTTCATAAACTCTGCTG 359  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 424 TGGTGTTACTGCAGCATGCGCTAGTTGGAGGTGCTAGCAACTGGAATTTCCAACCTGCCCA 483  
 QY 360 ACAATCACTCTGGGTATATCTCTGCTGATAACGGTTTGTGATTTGGATTTGGAATGTCC 419  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 484 ACAATAGCCTGGGCTTCATCTCTGGCAGATGCTGGTTTGGACGTGTGGATGGGAACAGCA 543  
 QY 420 GTGGAACGCTTTGGAGTAAAGGTTCATTCAACCTTTTCTCTTCATGATGAAGCTTTTCTGG 479  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 544 GGGGAACCCCTGGTCTCGAAACACAAACACACTCTCCATAGACCAGATGAGTTCTGGG 603  
 QY 480 ATTGGAGCTGGCAAGAGTGTAGCTGAATATGACCTTTTAGCAATGCTAGGCT---ATGTGT 536  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 604 CTTTCAGTTATGATGAGATGGCTAGCTTTCACCTTCTCGAGTGATAAACTTTATTTTGC 663  
 QY 537 ATACAGTCACAGATCCAAAATTTCTATATGTGGGGCATTCACAGGGAATATAATGGGTT 596  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 664 AGAAAACGGGCCAGGAAAAGATCTATTATGTGGCTATTTCACAGGGCACCACCATGGGCT 723  
 QY 597 TGGCGGCTTTGACGA 611  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 Db 724 TTATTGCATTTTCCA 738  
 ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||  
 RESULT 14  
 AAD19226  
 ID AAD19226 standard; DNA; 2035 BP.  
 XX AC AAD19226;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Human CG162 (or C59) lipase DNA #2.  
 XX KW Human; apolipoprotein; lipase; lipoprotein receptor; ALLr; angina;  
 KW cardiovascular disease; lipid metabolism; myocardial infarction;  
 KW cerebral ischaemia; arterial thrombosis; thrombolytic; antilipaeamic;  
 KW coronary artery thrombosis; cerebral artery thrombosis; stroke;  
 KW intracardiac thrombosis; gene therapy; cardiovascular; vasodilator;  
 KW neuroprotectant; cerebroprotective; ds.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 189..1418  
 FT /tag= a  
 FT /product= "Human CG162 (or C59) lipase protein"  
 FT sig\_peptide 189..245  
 FT /tag= b  
 FT mat\_peptide 246..1415  
 FT /tag= c  
 FT /product= "Human mature CG162 (or C59) lipase protein"  
 XX WO200179446-A2.  
 XX PD 25-OCT-2001.  
 XX PF 16-APR-2001; 2001WO-US12529.  
 XX PR 14-APR-2000; 2000US-197137P.  
 XX PR 20-JUN-2000; 2000US-0598042.  
 XX PR 03-AUG-2000; 2000US-0631451.  
 XX PR 22-SEP-2000; 2000US-0667298.  
 XX PR 17-NOV-2000; 2000US-0714936.  
 XX

(HYSE-) HYSEQ INC.  
XX Ballinger DG, Loeb D, Montgomery JR, Tang TY, Zhou P, Goodrich R;  
PI Liu C, Asundi V, Zhao QA, Wehrman T, Dmanac RT, Ren F, Qian XB;  
PI Wang D;  
XX WPI; 2001-611724/70.  
DR P-PSDB; AAE11931.  
XX Nucleic acids encoding human apolipoproteins, lipases, and lipoprotein  
PT receptor polypeptides, useful for preventing diagnosing and treating  
PT lipid metabolism disorders, thrombosis and cardiovascular diseases -  
XX  
PS Claim 1; Page 183-185; 266pp; English.  
XX The invention relates to polynucleotides encoding proteins CG122, CG179,  
CC CG95, CG121, CG162, CG27, CG153 and CG168 which are related to proteins  
CC involved in lipid metabolism and cardiovascular disease such as human  
CC apolipoproteins, lipases and lipoprotein receptor proteins. These DNA  
CC and protein sequences are useful for treating or preventing disorders  
CC associated with apolipoproteins, lipases and lipoprotein receptor (ALLr)  
CC expression and for treating lipid metabolism, cardiovascular diseases  
CC and thrombosis. Antibodies against these proteins are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of these sequences. ALLr polypeptides are also  
CC useful for identifying agents (agonists and antagonists) that bind to  
CC them and cells expressing ALLr proteins are useful for identifying a  
CC therapeutic agent for use in treatment of a pathology related to  
CC aberrant expression or physiological interactions of this polypeptide.  
CC Vectors comprising these DNA and protein sequences are also useful for  
CC producing ALLr proteins. The nucleic acids and polypeptides of the  
CC invention are also useful for the treatment of occlusive cardiovascular  
CC diseases, myocardial infarction, cerebral ischaemia, angina, arterial  
CC thrombosis, coronary artery thrombosis and cerebral artery thrombosis  
CC or intracardiac thrombosis and stroke. The nucleotides of the invention  
CC are used in gene therapy. The present sequence is human CG162 (or C59)  
CC lipase DNA.  
XX  
SQ Sequence 2035 BP; 608 A; 395 C; 456 G; 576 T; 0 other;  
Query Match 6.0%; Score 89.4; DB 22; Length 2035;  
Best Local Similarity 53.6%; Pred. No. 8.5e-14;  
Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;  
QY 183 TCCTGCACCTCGGCTACCCCTGCACGAGCACACAGCTTGAAACAAAGATGATCCTTT 242  
DB 304 TCCACATCAAGGCTATCCTCTGAGGATATGAAGTCGCACTGAAGATGGGTATATCC 363  
QY 243 TATCTCTTCAGCATATCCACATGGCAAAATAAAGCAGCAGATA---GTACTGGCCCTC 299  
DB 364 TTTCTGTTAACAGGATTCCTCGAGGCTAGTGCACCTAAGAAGACAGGTTCCAGGCCCTG 423  
QY 300 CAGTCTTTCTCAACATGCTCTTTTCAGGGAGGAGACACATGTTCTAAACTCTGCTG 359  
DB 424 TGTGTGTTACTGAGCATGGCCCTAGTTGGAGGTGCTAGCAACTGGATTTCACACTGCCCA 483  
QY 360 AGCAATCACTTGGGTATATCCTTGCTGTGATAACGGTTTGTGATTTGGATGGGAATGTC 419  
DB 484 ACAATAGCTGGCTTCATCTGGCAGATGCTGTTTGGACGTGGATGGGGAACAGCA 543  
QY 420 GTGGAACGGTTGGAGTAAGGTGATTCACACTTTCTTGTTCATGATAGCTTTTCTGGG 479  
DB 544 GGGAAACGCCCTGGTCTCGAAAAACAAGACACTCTCCATAGCAACCAAGATGATGTTCTGG 603  
QY 480 ATTGGAGCTGGCAAGATTTAGCTGATATGACCTTTTAGCAATGCTAGGCT---ATGTGT 536  
DB 604 CTTTCAGTTATGATGAGATGGGTAGGTTTGACCTTCTCSCAGCTGATAAACTTTATTTTCG 663  
QY 537 ATACAGTACACAGTCCAAATTTCTATATTTGGGGCATTCACAGGAACTAATATGGGTT 596  
DB 664 AGAAACGGGCCAGGAAAGATCTATTATGTCGGCTATTACAGGACCACCCATGGGCT 723  
QY 597 TGGCGGCTTTGACGA 611

DB 724 TTATTGCATTTTCCA 738  
RESULT 15  
AAF45131  
XX ID AAF45131 standard; cDNA; 2044 BP.  
XX AC AAF45131;  
XX DT 30-MAR-2001 (first entry)  
XX DE Human TANGO 294 cDNA.  
XX KW Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;  
KW central nervous system; focal brain disorder; bipolar affective disorder;  
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;  
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;  
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;  
XX neuropsychiatric; psychoactive substance use; anxiety; ss.  
OS Homo sapiens.  
XX WO200077239-A2.  
XX PD 21-DEC-2000.  
XX PF 24-MAY-2000; 2000WO-US14858.  
XX PR 14-JUN-1999; 99US-0333159.  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;  
XX WPI; 2001-032313/04.  
DR P-PSDB; AAB66065.  
XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for  
PT screening assays and diagnostic assays and for the treatment of  
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's  
XX disease -  
PS Claim 1; Fig 6; 359pp; English.  
XX The present invention relates to TANGO or INTERCEPT proteins and coding  
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,  
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding  
CC sequences are useful for the treatment of neurological disorders such as  
CC central nervous system (CNS) disorders, CNS-related disorders, focal  
CC brain disorders, global-diffuse cerebral disorders and other  
CC neurological and cerebrovascular disorders. The CNS disorders include  
CC Alzheimer's disease, senile dementia, Huntington's disease, amyotrophic  
CC lateral sclerosis, Parkinson's, Gilles de la Tourette's syndrome,  
CC autonomic function disorders such as hypertension and sleep disorders,  
CC neuropsychiatric disorders, psychoactive substance use disorders,  
CC anxiety, and bipolar affective disorder.  
XX  
SQ Sequence 2044 BP; 618 A; 401 C; 460 G; 565 T; 0 other;  
Query Match 6.0%; Score 89.4; DB 22; Length 2044;  
Best Local Similarity 53.6%; Pred. No. 8.5e-14;  
Matches 233; Conservative 0; Mismatches 196; Indels 6; Gaps 2;  
QY 183 TCCTGCCACTCGGCTACCCCTGCACGAGCACACAGTGTGAACAAAGATGATCCTTT 242  
DB 283 TCCACATCAAGGCTATCCTCTGAGGATATGAAGTCGCACTGAAGATGGGTATATCC 342  
QY 243 TATCTCTTCAGCATATCCACATGGCAAAATAAAGCAGCAGATA---GTACTGGCCCTC 299  
DB 343 TTTCTGTTAACAGGATTTCTCGAGGCCCTAGTGCACACTAAGAAGACAGGTTCCAGGCCCTG 402  
QY 300 CAGTCTTTCTCAACATGCTCTTTTTCAGGGAGGAGACACATGTTCTAAACTCTGCTG 359

[illegible]

Search completed: December 30, 2002, 17:23:18  
Job time : 291 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2002, 17:13:26 ; Search time 2226 Seconds  
(without alignments)  
10789.716 Million cell updates/sec

Title: US-09-699-652A-13

Perfect score: 1483

Sequence: 1 gcacgagtcacagcgggc.....ctctcaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549.6	37.1	717	10 BE443524	BE443524 WHE1115_C
2	464.6	31.3	623	13 BM324154	BM324154 PIC1_24_G
3	403	27.2	668	13 BI075779	BI075779 IP1_21_B0
4	360	24.3	491	10 BE367129	BE367129 PIC1_43_D1
5	357.4	24.1	478	9 AU088737	AU088737 AU088737
6	341.6	23.0	631	13 BM037885	BM037885 S113f11 S

7	327.6	22.1	828	14	BO122157	BO122157
8	327.4	22.1	441	13	BM501471	BM501471 PAC000000
9	324.2	21.9	365	9	AU161807	AU161807 AU161807
10	324.2	21.9	572	13	BM326139	BM326139 PIC1_56_B
11	316.4	21.3	476	10	BE367231	BE367231 PIC1_43_D1
12	294.4	19.9	647	13	BM328649	BM328649 PIC1_24_G
13	285.8	19.3	323	9	AU161802	AU161802 AU161802
14	285.2	19.2	646	13	BI075485	BI075485 IP1_21_B0
15	281	18.9	521	9	AU088738	AU088738 AU088738
16	280.4	18.9	420	9	AJ474806	AJ474806 AJ474806
17	277.6	18.7	449	9	AU174368	AU174368 AU174368
18	263.8	17.8	635	14	BO122609	BO122609 EST608185
19	263.2	17.7	600	10	BE191250	BE191250 sn94a09.y
20	253	17.1	659	13	BI405491	BI405491 13BE11 Ma
21	245.6	16.6	588	13	BM331443	BM331443 PIC1_56_B
22	240.4	16.2	549	13	BM527928	BM527928 sal52810
23	231.2	15.6	394	10	BE356956	BE356956 DGI_145_F
24	227.2	15.3	501	13	BI784642	BI784642 sal57e04
25	225.2	15.2	492	9	AL380410	AL380410 MBB52C08
26	215.2	14.5	422	14	BO756230	BO756230 EB6m06_SQ
27	213.6	14.4	536	13	BI321617	BI321617 sal15e01
28	212	14.3	684	14	BO407284	BO407284 GA_E010
29	210.6	14.2	630	14	BO415924	BO415924 GA_E010
30	209.6	14.1	600	10	AV559339	AV559339 AV559339
31	204.6	13.8	563	10	AW277778	AW277778 sf86e03.y
32	204.6	13.8	625	12	BE822354	BE822354 GM700017A
33	203	13.7	702	12	BF479168	BF479168 L48-2822T
34	202.2	13.6	469	12	BF325456	BF325456 su33a06.y
35	201.4	13.6	650	14	BO124957	BO124957 EST610533
36	199.4	13.4	591	14	BO743009	BO743009 sqs58q09
37	197.2	13.3	624	12	BE822717	BE822717 GM700018A
38	195.6	13.2	329	12	BG313685	BG313685 WHE2057_C
39	195.2	13.2	461	12	BE800757	BE800757 sq97h02.y
40	193.6	13.1	558	13	BI427293	BI427293 sab78e08
41	190.4	12.8	416	10	AW830986	AW830986 sm31d03.y
42	189.6	12.8	573	10	BE318901	BE318901 NF004D05L
43	188.8	12.7	514	14	BO122585	BO122585 EST608161
44	185.6	12.5	446	12	BF716146	BF716146 saal5G06
45	185.4	12.5	340	9	AU223277	AU223277 AU223277

## ALIGNMENTS

RESULT 1

BE443524

LOCUS

DEFINITION

WHE1115\_C05\_F09ZS Wheat etiolated seedling root normalized cdna

Library Triticum aestivum cdna clone WHE1115\_C05\_F09, mRNA

sequence..

ACCESSION

BE443524

VERSION

BE443524.1

KEYWORDS

EST..

SOURCE

bread wheat.

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

REFERENCE

1 (bases 1 to 717)

AUTHORS

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han

, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,

Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cdna library

Unpublished (2000)

JOURNAL

Contact: Olin Anderson

COMMENT

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low



Best Local Similarity 84.1%; Pred. No. 6.4e-98;  
Matches 524; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
QY 267 GCAAAATAAACACAGACTACTGCGCCCTCCAGTTTCTTCAACATGCTCTTTTC 326
Db 1 GCACGAGGGAATACAGATAATGCTGGACCTCCAGTTTTCTTCAGCATGCTCTTTTC 60

QY 327 AGGGAGGAGACACATGGTTTCATAAACTCTGCTGAGCAATCACTTGGGTATATCTTTGCTG 386
Db 61 AGGGTGGAGATACATGGTTTCATAAACTCAATGAACAATCACTTGGATATATCTTTGCTG 120

QY 387 ATACCGTTTTGATGTTGGATGGGAATGCTCCGCGGAACGCGTTGGAGTAAAGGTCAAT 446
Db 121 ATAATGGTTTTGATGTTGGATGGGAATGCTCCGCGGACACGTTGGAGTAAAGGCCACT 180

QY 447 CAACCTTTCTGTCATGATAAGCTTTCTGCGGATTGGAGCTGGCAAGAGTTAGTCTGAAT 506
Db 181 CTACTCTCTGTCATGATAAGCTTTCTGCGAATGGATTGGGAACCTTGGCTGAAT 240

QY 507 ATGACCTTTTAGCAATGCTAGGCTATGTATACAGTACACAGTCCAAAATCTATATG 566
Db 241 ACGATGTTTTGGCAATGTTAAAGCTATGTATATACAAATACACAGTCCAAAATCTATATG 300

QY 567 TGGGGGATTCACAGGAACATAAATGGGTTTGGCGGCTTTGACGATGCCGGAATAGTAA 626
Db 301 TGGGACATTCACAGGGAACATATCATGGGTCTGGCTGCGTTTACAATGCCCTGAATAGTAA 360

QY 627 AAATGATTAGCTCTCAGACACTCTTTGTCCTATTCTTATCTTATCTTATCATGCTAGTGCTA 586
Db 361 AAATGATAAGCTCTGCTGTCCTTTGTCCTATTCTTATCTTATCTTATCATGCTAGTGCTA 420

QY 687 GTTTTGTCTCAGAGCAGTCGCATGTCATCTTGATCAGATGCTTGTACTATGGGAATTC 746
Db 421 GTTTTGTCTTAGACAGTGTCCATGCACTTTGACCAATGCTTGTGCTATGGGCATCC 480

QY 747 ACCAGTGAACCTTCGTAGCGACATAGGGGGTTCAAAATAGTAGATTCTTTGTCGATGGTG 806
Db 481 ATCAGTTGAACCTTCGGAGTGATATGGGTGTCAGATTTTAGATTCTGCTGTGATGATG 540

QY 807 AACAGTGTAGTGAACAATTTGCTATCTGCGATTACAGGGGAAAACCTTCTCTTCAATA 866
Db 541 ACATTTGAGCTGCACAGATCTGTTATCTTCAATACAGGTCAAAACCTGTTGTTCAAT 600

QY 867 CATCAGGATGATTATTTTGG 889
Db 601 CATCTCGGATTGACTATTATTG 623
```

RESULT 3  
BI075779  
LOCUS  
DEFINITION  
IP1\_21\_B03.bl\_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BI075779  
BI075779.1 GI:14514436  
EST.  
sorghum.  
Sorghum bicolor

REFERENCE  
AUTHORS  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 668)  
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt  
,L.H.

TITLE  
JOURNAL  
COMMENT  
An EST database from Sorghum: developing preanthesis pannicles  
Unpublished (2001)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
below Phred quality 16. The threshold for high quality sequence is  
20. Three-prime sequences, which are obtained with PolyMix or T7  
sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV  
High quality sequence stop: 572  
POLYA-No.

FEATURES  
Source

Location/Qualifiers

1. 668  
/organism="Sorghum bicolor"  
/cultivar="Brx623"  
/db\_xref="taxon:4558"  
/clone\_lib="Immature pannicle 1 (IP1)"  
/note="Organ: Developing preanthesis pannicles; Vector:  
pBluescript II SK(-) from Lambda Zap II; Site: 1. XhoI;  
Site: 2. EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."

BASE COUNT 145 a 176 c 175 g 171 t 1 others

Query Match 27.2%; Score 403; DB 13; Length 668;  
Best Local Similarity 78.9%; Pred. No. 1.5e-83;  
Matches 494; Conservative 0; Mismatches 126; Indels 6; Gaps 1;

```
QY 23 GCGTTGGCGATGGCGGGCCACGCCCGGAGGAGCGCTCCCTCTGATCTCTCTC 82
Db 42 GCGGTGGCGATGGTGGGCCAGGAGGCGTCCGCCGGGCCCGCGCTCTCTGTGCTC 101

QY 83 GTGCTCTCTGCTGGGTGCGATCTCTCCGAGAGCTCCCAAGCGCGCGCCCTCCGC 142
Db 102 GTGCTCTCTCTCTCTCGCGGTGAGCGCCGCGATCCCGGCCACAGACGCTCGC 161

QY 143 CGCGTGGGTCC-----GGCTCCGGCGCTCTCGGACCAGCTGCTCTGCCACTCGC 196
Db 162 CGCGTCTCCCGCTCGCGCGCGCGCGCTCTGCCAGCAGCTGCTCTGCCGAGGCG 221

QY 197 TACCCCTGACCGACACACCGTTGAAACAAAAGATGATCTTTTATCTTTCAGCAT 256
Db 222 TACCGTGCACCGACACACCGTTCAACGGATGATGCTCTTTGCTCTTCAGCAT 281

QY 257 ATCCACATGGCAAAATAAAGCAGCAGATAGTACTGGCCCTCCAGTTTTCCTCAACAT 316
Db 282 ATTCACATGGCAAAATGGAATACAGATATGCTGGACCTCCAGTTTTCCTCAGCAT 341

QY 317 GGTCTTTTCAGGAGGAGACACATGTTTCATAAACTCTGCTGACCAATCACTTGGGTAT 376
Db 342 GGTCTTTTCAGGAGGAGATACATGGTTTCATAAACTCCAATGAACAATCACTTGGATAT 401

QY 377 ATCCTTGTGATAACCGTTTTCATGTTGGATTGGGAATGTCCTGGAAACGCGTTGGAGT 436
Db 402 ATCCTTGTGATAACCGTTTTCATGTTGGATTGGGAATGTCCTGGAAACGCGTTGGAGT 461

QY 437 AAAGTCATTTCAACCTTTTCTGTCATGATAAGCTTTTCTGGGATTGGAGCTGGCAAGAG 496
Db 462 AAAGCCACTCTACTCTCTCTCTCATGATAAGCTTTTCTGGGAATGGAGTTGGCAAGAG 521

QY 497 TTAGCTGAATATGACCTTTTAGCAATGCTAGCTATGTATATACAGTACACAGTCCAAA 556
Db 522 CTGTGTAATAGTAGTTTGTGGCAATGTTAAGCTATGATATATACAAATTACAGTCCAGA 581

QY 557 ATTCTATATGTGGGCATTTCACAGGGAACATAAATGGTTTGGCGCTTTTGACGATGCC 616
Db 582 ATTCTATATGTGGGACATTTCACAGGGAACATAATCTGGGCTCTGGCTTTACNATGCT 641

QY 617 GAAATAGTAAAAATGATTAGCTCTGC 642
Db 642 GAAATAGTAAAAATGATAAGCTCTGC 667
```

RESULT 4  
BE367129  
LOCUS

BE367129 491 bp mRNA linear EST 20-JUL-2000

DEFINITION P11\_43\_D12.bl\_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION BE367129  
VERSION BE367129.1 GI:9308686  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 491)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt  
L.H.

TITLE An EST database from Sorghum: pathogen-induced plants

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV

High quality sequence stop: 458

POLYA=No.

FEATURES Location/Qualifiers

source 1..491

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Pathogen induced 1 (P11)"

/note="Organ: Anthracnose-infected leaves from

two-week-old sorghum plants 48 hr after inoculation;

Vector: pBluescript II from Lambda Zap II; Site\_1: XhoI;

Site\_2: EcoRI; Two-week-old sorghum plants (BTX 623

cultivar) were infected with pathogen (isolate FRM421 of

Colletotrichum graminicola, which is a sorghum isolate).

RNA was prepared from infected leaves harvested from 45

seedlings 48 hours after inoculation. Note: young

seedlings (2 weeks old) exhibit juvenile resistant

reaction, which is an incompatible interaction. As they

grow older (4 weeks or older), plants resume susceptibility

to anthracnose disease. The library was made from poly-A

RNA in the cloning vector Lambda Zap II. Clones to be

sequenced were prepared by mass excision. WARNING: While

most or all ESTs are expected to derive from the host

plant, no effort was made to eliminate ESTs deriving from

the pathogen."

BASE COUNT 131 a 104 c 122 g 134 t

ORIGIN

Query Match 24.3%; Score 360; DB 10; Length 491;

Best Local Similarity 83.6%; Pred. No. 1.7e-73;

Matches 408; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 706 CGCCATGATCTTGATCAGATCGTTGTACTATGGGATTCACGACGTGAACCTCCGTAG 765

DB 4 CGAGGTGATCTTGACAGATCGTTGTGCTATGGGCATCCATCAGTTGAACCTCCGGAG 63

QY 766 CGACATGGGGTTCAATAGTAGATCTTTGTGCGATGGTGAACACGTGGATTGCAACAA 825

DB 64 TCATATGGGTGTCAGATTTTATAGTCGCTGTGATGATGACATTTGGAGTGAACGA 123

QY 826 TTTGCTATCTCGGATTACAGGGGAAACATGTTGCTTCAATACATCAAGATGATTATTA 885

DB 124 TCTGTTATCTTCAATAACAGGTCAAACATGTTGTTTCAATTCATTCGAGTATTA 183

QY 886 TTTGGAGTATGAACCTCATCCATCATCACAACAAATCTGCACATCTTTTCAGATGAT 945

DB 184 TTTGGAGTATGAACCCCATCCGCTCATCAACAAAAATTTGGGCGATCTTTTCCAGATGAT 243

QY 946 CAGGAAAGGCACCTTCCCAAGTATGACTATGGTATTGGGAAACCTAAGCGCTACGG 1005  
DB 244 CAGCAAGGCAGTTTCGCAAGTATGACTATGGTATGGGGAAACATAAGGCACCTACGG 303  
QY 1006 TCATTTGGTCTCTCCCGCATTTTGACCTTAAGCAGCATACCAGAATCACTGCCCATATGGAT 1065  
DB 304 CCACGGGCATCTCTCCCTCATTCGATTAAGCAGCATACCGGAATCACTGCCAATATGAT 363  
QY 1066 GGGATATGGAGGCTCTGATGCATTGGCTGATTAACCGATGTTCCAGCGTACTACTACAGA 1125  
DB 364 GGGTATGGAGGCTCTGATGCATGGCTGACGTAAACAGATGTTGAGCGCACCATCAAGA 423  
QY 1126 GCTGGGATCTACACCAAGAACTTCTGTACATTTGGTACTATGGCCATATGATTTTGTAT 1185  
DB 424 GTTGAGATCCACGCCAGAGCTGCTGTACATTTGGTACTATGGCCACATTTGATTCATCAT 483  
QY 1186 GAGCGTGA 1193  
DB 484 GAGCGTGA 491

RESULT 5

AU088737

LOCUS

DEFINITION AU088737 Rice shoot Oryza sativa (japonica cultivar-group) cDNA

clone S5305, mRNA sequence.

ACCESSION AU088737

VERSION AU088737.1 GI:7378466

KEYWORDS EST.

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group).

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartioideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 478)

AUTHORS

Sasaki,T. and Yamamoto,K.

TITLE

Rice cDNA from etiolated shoot (2000)

JOURNAL

Unpublished (2000)

COMMENT

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@nri.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/

PROJECT = 'RGP'

FEATURES Location/Qualifiers

source 1..478

/organism="Oryza sativa (japonica cultivar-group)"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="S5305"

/clone\_lib="Rice shoot"

/dev\_stage="Etiolated shoot (8 days old)"

/note="Etiolated shoot (8 days old)"

BASE COUNT 98 a 148 c 120 g 110 t 2 others

ORIGIN

Query Match 24.1%; Score 357.4; DB 9; Length 478;

Best Local Similarity 94.6%; Pred. No. 6.8e-73;

Matches 434; Conservative 0; Mismatches 18; Indels 7; Gaps 6;

QY 10 CACAGCGCGCGGCGTTGGCGATGGCGATCGCGGCCACGCCCGGAGGAGCGCTCCC 69

DB 24 CAAGCGCGCGGCGTTGGCGATGGCGATCGCGGCCACG-CCCCGAGGAGCGCTCCC 82

QY 70 CTGATCTCTCTCGTCGTCCTTGTGTCGGTCCGATCGTCCTCCGAGCCTCCCCAGCGCG 129

DB 83 CTGATCTCTCTCGTCGTCCTTGTGTCGNGTCGATCGTCCTCCGAGCCTCCCCAGCGCG 142

QY 130 CGCGCCCTCCCGCGCTCGGCTCCGGCTCCGGCGGCTCTTCGG-ACCAGCTGCTCTGCG 188

DB 143 CGCGCCCTCCCGCGCTCGGCTCCGGCTCCGGCGGCTCTTCGGAACCAAGCTGCTCTGCG 202

Qy	939	A-GATGATCAGGAAGGCACTTTCGCAAA-----AGTATGACTATGGGTATTGGGAACCC	992
Db	186	AGGAGGATTCAGAAAGGCACATTCGCCAAAGTATGACTATTTGGGGTTATTGGGGAACCC	245
Qy	993	TAAGCGCGTACGGTCATTTTGGG-TCCTCCCGCATTTTGACC--TAACGAGCATACCCAGAA	1048
Db	246	TAAGCGCGTACGGTCATNTGGCTTCCTCCCGCAATTTTGACCCTTAAGCAGCATACCCAGAA	305
Qy	1049	TCACTGCCCATATGATGGGATATGGAGGCTCTTGATGCATTTGGCTGATGTAAACCGATGTT	1108
Db	306	TCACGGCCCATATGATGGGATATGGAGGCTCTTGATGCATTTGGCTGATGTAAACCGATGTT	365
Qy	1109	CAGCGTACTATCAGAGAGCTGGGATCTAC-ACCAGAACCTCTCTGACATTTGGTGTACTATGG	1167
Db	366	CAGCGTACTATCAGAGAGCTGGGATCTACAAACCCAGAACTTCTGTACATTTGGTGTACTATGG	425
Qy	1168	CCATATTTGATTTTGTATGACGCGTGAAGGCGAAAGATGATGCTTTTATGTGGACCTTAATAAG	1227
Db	426	CCATATTTGATTTTGTATGACGCGTGAAGGCGAAGATGATGCTTTTATGTGGACCTTAATAAG	485
Qy	1228	ATTTCTTAGGGAAATGGATGGCATAATAGCTATTAGGATGCTCTTCATGTGTATAATAAA	1287
Db	486	ATTTCTTAGGGAAATGGATGGCATAATAGCTATTAGGATGCTCTTCATGTGTATAATAAA	545
Qy	1288	AACATCTGTACAGTATTTGGTCTCTCCGGATGTGAGTATGTATATATTTGCAATATGAGCTT	1347
Db	546	AACATCTGTACAGTATTTGGTCTCTCCGGATGTGAGTATGTATATATTTGCAATATGAGCTT	605
Qy	1348	GTTCGATCTA 1357	
Db	606	GTTCGATCTA 615	
RESULT 7			
BO122157			
LOCUS	EST607733	GLSD Medicago truncatula cDNA clone pGLSD-28123, mRNA	EST 17-APR-2002
DEFINITION	BO122157	828 bp mRNA linear	
ACCESSION	BO122157		
VERSION	BO122157.1	GI:20174119	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
REFERENCE	1 (bases 1 to 828)		
AUTHORS	Grusak,M.A., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cheung		
	,F. and Fraser,C.M.		
TITLE	ESTs from late stage developing seeds of Medicago truncatula		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Grusak, M.A. USDA/ARS Children's Nutrition Research Center Baylor College of Medicine 1100 Bates Street, Houston, TX 77030-2600, USA Tel: 713 798 7044 Fax: 713 798 7078 Email: mgrusak@bcm.tmc.edu TIGR sequence name: WTRAA60TK More information is available at: www.medicago.org Seq primer: SKmod (CTA gAA CTA gTg gAT CC).		
FEATURES	Location/Qualifiers		
source	1..828		
	/organism="Medicago truncatula"		
	/cultivar="A17"		
	/db_xref="taxon:3880"		
	/clone="pGLSD-28123"		
	/clone_lib="GLSD"		
	/tissue_type="Immature seeds"		
	/dev_stage="25 to 35 days after pollination"		
	/lab_host="XLOLR"		
	/note="Vector: pBluescript SK-; Site_1: EcorI; Site_2:		





```
QY 377 ATCCTTGCTGATAACGGTTTGGATTTGGATGGGAATGCCGTGGAAACGCGTTGGAGT 436
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 406 ATCCTTGCTGATAAATGGTTTGGATTTGGATGGGAATGCCGTGGCAACGCTGGAGT 465
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 437 AAAGTCATTCAACCTTTCTGTTTCATGATAGCTTTTCTGGGATTTGGAGCTGGCAAGAG 496
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 466 AAAGGCCACTCTACTCTCTCTGTTTCATGATAAGCTTTTCTGGGATTTGGAGTGGCAAGAC 525
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 TTAGCTGAATAGACCTTTTACCAATGCTAGGCTATGTTGATACAGT 543
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 CTGCTGAATACGATGTTTGGCAATGTTAAGCTATGATATACAAT 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
BE367231
LOCUS BE367231 476 bp mRNA linear EST 20-JUL-2000
DEFINITION P11_43_D12.g2_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION BE367231
VERSION BE367231.1 GI:9308788
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 476)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 13
High quality sequence stop: 473
POLYA=No.
FEATURES
source
1..476
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate FRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."
BASE COUNT 133 a 97 c 115 g 131 t
ORIGIN
Query Match 21.3%; Score 316.4; DB 10; Length 476;
Best Local Similarity 83.4%; Pred. No. 2.5e-63;
Matches 372; Conservative 0; Mismatches 71; Indels 3; Gaps 1;
```

```
QY 826 TTTGCTATCTCGGATTACAGGGGAAACCTGTTGCTTCAATACATCAAGGATTGATTATTA 885
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TCTGTTATCTCATACACAGGTCAAAACTGTTGTTCAATTCACTCGGATTGACTATTA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 886 TTTGGAGTGAACCTCATCCATCATCGACAAAAAATCTGCACCATCTTTTTCAGATGAT 945
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TTTGGAGTGAACCCCATCCGTCATCAACAAAAAATTTGCGGCATCTTTTCCAGATGAT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 946 CAGAAAGGCACCTTTCGCAAGATGATGACTATGGTTATTGGGAAACCTAAGGCCCTACGG 1005
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CAGAAAGGCAGTTTCGCAAGATGATGACTATGGTTATTGGGAAACCTAAGGCCCTACGG 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1006 TCATTTGCTCTCCCGCATTTGACCTAAGCAGCATCACCAATCACTGCCCATATGGAT 1065
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CCAGCGCATCTCCCTCATTCGATCTAAGCAGCATACCGGAATCACTGCCAATATGGAT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1066 GGGATATGGAGGCTTGTGATGCAATGGCTGATTAACCGATGTTTCAGCGTACTATCAGAGA 1125
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GGGTATGGAGGCTTGTGATGCACTGGCTGACGTAAACAGTGTTCAGCGCACCATCAAGA 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1126 GCTGGATCTACACAGAACCTTCTACATGCTGACTATGGCCATATGATTTTGTAT 1185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GTTGAGATCCACGCGCAGAGCTGCTACATTTGGTACTATGGCCACATTTGATTTTCATCAT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1186 GAGCGTGAAGCGGAAGATGATGTTTATGCGACCTAATAAGATTTCTTAGGG---AAAA 1242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GAGCGTGAAGCGGAGGACGATGTTTATGTTGACTTAAATGAGGTTTCTCAGGGGCCACCA 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1243 TGGATGCGATAATAGCTATTAGGATG 1268
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GGAATGCACAGTAGTTATTAGGACG 446
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
BM328649
LOCUS BM328649 647 bp mRNA linear EST 04-JAN-2002
DEFINITION P1C1_24_G11-gl_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
bicolor cDNA, mRNA sequence.
ACCESSION BM328649
VERSION BM328649.1 GI:18067786
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 647)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: plants infected with a compatible
pathogen
JOURNAL Unpublished (2002)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector, and regions
below Phred quality 16. The threshold for highest quality sequence
is 20. Three-prime sequences, which are obtained with PolyTMix or
T7 sequencing primer, are presented as the reverse complement.
Seq primer: 17
High quality sequence start: 61
High quality sequence stop: 647
POLYA=Yes.
FEATURES
source
1..647
/organism="Sorghum bicolor"
/cultivar="Btx623"
/db_xref="taxon:4558"
```



/clone\_lib="Pathogen-infected compatible 1 (PIC1)"  
/tissue\_type="Leaves"  
/dev\_stage="4-week-old seedlings infected with  
Colletotrichum graminicola"  
/note="vector: pBluescript II SK(-) from Lambda Zap II;  
Site 1: XhoI; Site 2: EcoRI; Four-week-old sorghum  
seedlings were sprayed with spore suspension prepared from  
3-week-old FRM421, a sorghum isolate of the anthracnose  
pathogen Colletotrichum graminicola. Inoculated plants  
were kept in a 25 C dark growth chamber with 100% relative  
humidity for 24 hr, followed by 12/12 hr of light/dark  
cycle at 25 C with 90% relative humidity for another 24  
hr. All leaves were harvested and quick frozen with liquid  
nitrogen and stored in a -80 C freezer. The library was  
made from poly-A RNA in the cloning vector lambda ZAP II.  
Clones to be sequenced were prepared by mass excision.  
WARNING: While most or all ESTs are expected to derive  
from the host plant, no effort was made to eliminate ESTs  
deriving from the pathogen."  
BASE COUNT 182 a 128 c 158 g 179 t  
ORIGIN

Query Match 19.9%; Score 294.4; DB 13; Length 647;  
Best Local Similarity 78.7%; Pred. No. 3.5e-58;  
Matches 365; Conservative 0; Mismatches 96; Indels 3; Gaps 1;  
Qy 873 GGATTGATTATTATTGGAGTATGAACCTCATCCATCATCGACAAAATCTGCACCATC 932  
Db 2 GGATTGACTATTATTGGAGTATGAACCCATCGTCATCAACAAAATTTGGCGCATC 61  
Qy 933 TTTTTCAGATGATCAGGAAGGCACTTTTCGCAAAAGTATGACTATGGGTATTGGGAAC 992  
Db 62 TTTTCCAGATGATCAGGAAGGCAAGTTTCGCAAAAGTATGACTATGGATGGTGGGAAACA 121  
Qy 993 TAAGCGCTACGGTCATTTCGCTCCTCCCGCATTTGACCTTAAGCAGCATACCAAGATCAC 1052  
Db 122 TAAGCGCTACGGTCATTTCGCTCCTCCCGCATTTGACCTTAAGCAGCATACCGGAATCAC 181  
Qy 1053 TGCCCATATGATGGATATGAGGCTTTGATGCTTGGCTGATGTAACCGATGTTACG 1112  
Db 182 TGCCAATATGATGGGTATGAGGCTTTGATGCTTGGCTGATGTAACCGATGTTGAGC 241  
Qy 1113 GTACTATCAGAGCTGGGATCTACACCAAGCTTCGTACATGGTGACTATGGCCATA 1172  
Db 242 TCACCATCAAGAGTGTAGATTCACGCCAGAGCTGCTGTACATGGTGACTATGGCCACA 301  
Qy 1173 TTGATTTTGTATGAGCTGAAGCGGAAGATGATGTTTATGTGACCTTAATACATTC 1232  
Db 302 TTGATTTTGTATGAGCTGAAGCGGAAGATGATGTTTATGTGACCTTAATACATTC 361  
Qy 1233 TTAGG---AAAATGGATGGCATATAGCTATTAGGATGCTTTCATGTGTATAATAAAA 1289  
Db 362 TCAGGCCCCAGCAGGAATGCACAGTAGTTATTAGGAGCATCGTATGCCCTTAATTAA 421  
Qy 1290 CATCTGTACATATTGGTCTCTCCCGATGAGTATGTATATA 1333  
Db 422 TGTATTACTATGATGACGATACGACTATGTAATACGTACATA 465

RESULT 13  
LOCUS AU161802 323 bp mRNA linear EST 03-APR-2002  
DEFINITION AU161802 Rice shoot Oryza sativa (japonica cultivar-group) cDNA  
clone S5639, mRNA sequence.  
ACCESSION AU161802  
VERSION AU161802.1 GI:10931528  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group).  
ORGANISM Oryza sativa (japonica cultivar-group).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 323)

AUTHORS Sasaki, T. and Yamamoto, K.  
TITLE Rice cDNA from etiolated shoot (2000)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@nri.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
PROJECT "RGP"  
Location/Qualifiers  
1. 323  
/organism="Oryza sativa (japonica cultivar-group)"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="S5639"  
/clone\_lib="Rice shoot"  
/dev\_stage="Etiolated shoot (8 days old)"  
/note="Etiolated shoot (8 days old)"  
BASE COUNT 59 a 119 c 83 g 61 t  
ORIGIN 1 others

Query Match 19.3%; Score 285.8; DB 9; Length 323;  
Best Local Similarity 99.0%; Pred. No. 3.5e-56;  
Matches 298; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
Qy 10 CACAGCGCGCGGGCTTGGCGATGGCGATGGCGGCCACGCCCGGAGGAGCGCTCCC 69  
Db 23 CAAAGCGCGCGGGCTTGGCGATGGCGGCCACG-CCCCGGAGGAGGCGCTCCC 81  
Qy 70 CCTGATCTCTCTCGTCTCTTGTCTGGGTGCGATCGTCTCGGAGCTCCCGAGCGC 129  
Db 82 CCTGATCTCTCTCGTCTCTTGTCTGGGTGCGATCGTCTCGGAGCTCCCGAGCGC 141  
Qy 130 CGCGCCCTCCCGCGGTGCGGTCCGGCTCCGGCGGCTCTCGGAGCTCCCGAGCGC 189  
Db 142 CGCGCCCTCCCGCGGTGCGGTCCGGCTCCGGCGGCTCTCGGAGCTCCCGAGCGC 201  
Qy 190 ACTCGGCTACCCCTCGACCGAGCACAGCTTGAACAAAGATGATTCCTTCTCT 249  
Db 202 ACTCGGCTACCCCTCGACCGAGCACAGCTTGAACAAAGATGATTCCTTCTCT 261  
Qy 250 TCAGCATATCCCATGATGCAAAAATAAAGCAGCAGATAGTACTGGCCCTCCAGTTTTTCT 309  
Db 262 TCAGCATATCCCATGATGCAAAAATAAAGCAGCAGATAGTACTGGCCCTCCAGTTTTTCT 321  
Qy 310 T 310  
Db 322 T 322

RESULT 14  
LOCUS BI075485 646 bp mRNA linear EST 20-JUN-2001  
DEFINITION IP1\_21\_B03\_g1\_A002 Immature pannicle 1 (IPI) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION BI075485  
VERSION BI075485.1 GI:14514142  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 646)  
AUTHORS Klein, R.R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and Pratt, L.H.  
TITLE An EST database from Sorghum: developing preanthesis pannicles  
JOURNAL Unpublished (2001)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below three quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement.  
Seq primer: T7  
High quality sequence start: 12  
High quality sequence stop: 646  
POLYA=tes.

#### FEATURES

source

Location/Qualifiers  
1..646  
/organism="Sorghum bicolor"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone\_lib="Immature pannicle 1 (IP1)"  
/note="Organ: Developing preanthesis pannicles; Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
182 a 132 c 156 g 176 t

BASE COUNT

ORIGIN

Query Match 19.2%; Score 285.2; DB 13; Length 646;  
Best Local Similarity 77.9%; Pred. No. 4.9e-56;  
Matches 357; Conservative 0; Mismatches 98; Indels 3; Gaps 1;  
QY 879 ATTATTATTGGAGTATGAACCTTCATCCATCGACAAAAATCTGCACCATCTTTTTC 938  
Db 1 ACTATTATTGGAGTATGAACCCCATCCGTCATCAACAAAAATTTGGCGCATCTTTTC 60  
QY 939 AGATGATCAGGAAGGCACCTTTCCGAAAGTATGACTATGGTTATTGGGAAACCTAAGGC 998  
Db 61 AGATGATCAGGAAGGCAGTTTCGCAAGTATGACTATGGTGGGGAAACATAGGC 120  
QY 999 GCTACGGTCATTTCGCTCCCGCATTTGACCTAAGCAGCATACCAAGATCACTGCCCCA 1058  
Db 121 ACTACGCCAGCGCATCTCCCTCATTCATCTAAGCAGCATACCGAATCACTGCCAA 180  
QY 1059 TATGGATGGGATATGGAGGCTTGTATGCAATGGTGTATGTAACCGATGTTTCACGGCTACTA 1118  
Db 181 TATGGATGGGATATGGAGGCTTGTATGCACTGGCTGACGTAAACAGATGTTTGAGCGCACCA 240  
QY 1119 TCAGAGCTGGGATCTACACCAAGCTTCTGTACATTTGTGACTATGGCCATATTGATT 1178  
Db 241 TCAAAGAGTTGAGATCCACGCCAGAGCTGTGTACATTTGTGACTATGGCCACATTTGATT 300  
QY 1179 TTGTTATGAGCGTGAAGCGGAAGATGATCTTTATGTGGACCTAAATAAGATTTCTTAGGG 1238  
Db 301 TCATCATGAGCGTGAAGCGGAGGACGATGTTATGTGACTCTATGAGGTTCTCAGGG 360  
QY 1239 ---AAAATGGATGGCATAATAGCTATTAGGATGCTTTTCATGTGTATTAATAAAACACTGT 1295  
Db 361 CCCAGCAGGGAATCCACAGTAGTTATTAGGAGCGCATCGTATGCCCTTAAATTAATGTTTA 420  
QY 1296 TACAGTATTGGTCTCTCCCGATGTGAGTATGATATA 1333  
Db 421 TACTATGATGACAGTACGACTATGTAATAACGTTACATA 458

RESULT 15

LOCUS

AU088738 521 bp mRNA linear EST 02-APR-2002  
DEFINITION AU088738 Rice shoot Oryza sativa (japonica cultivar-group) cDNA clone S5305, mRNA sequence.

ACCESSION AU088738

VERSION AU088738.1 GI:7378467

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group).

ORGANISM Oryza sativa (japonica cultivar-group).

#### REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 521)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from etiolated shoot (2000)  
Unpublished (2000)  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@r.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
PROJECT = 'RGP'.

#### FEATURES

source

Location/Qualifiers

1..521

/organism="Oryza sativa (japonica cultivar-group)"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone\_lib="S5305"

/dev\_stage="Rice shoot"

/note="Etiolated shoot (8 days old)"

/note="Etiolated shoot (8 days old)"

149 a 111 c 100 g 160 t 1 others

BASE COUNT

ORIGIN

Query Match 18.9%; Score 281; DB 9; Length 521;  
Best Local Similarity 90.7%; Pred. No. 4.7e-55;  
Matches 332; Conservative 0; Mismatches 31; Indels 3; Gaps 3;  
QY 174 ACCAGTGTCTCGCCACTCGGCTACCCCTGCGCAGGACACACGTTGAAACAAAAGATG 233  
Db 7 AACAGCTNGTCTCCAACTCGGTAAGCTGCACGGAG-TCACAACGTTGAAACAAAAGATG 65  
QY 234 GATTCCTTTTATCTCTTCAGCATATCCACATGCGCAAAATAAAGCAGCAGATAGTACTG 293  
Db 66 GATTCCTCTAACTCTTCAGCATATTCACATGGCAAAATAAACCAGCAGATAGTACT- 124  
QY 294 GCCTCCAGTTTTTCTTCAACATGGTCTTTTTCAGGGAGGAGACACATGTTTCATAAACT 353  
Db 125 GCCTCCAGTTTTTCTTCAACAT-GTCTTTTTCAGGGAGGAGACACATGTTTCATAAACT 183  
QY 354 CTGCTGAGCAATCATTGGGTATATCTTCTGCTGATACGGTTTTTGTGATTGGGGA 413  
Db 184 CTGCTGAGCAATCATTGGGTATATCTTCTGCTGATACGGTTTTTGTGATTGGGGA 243  
QY 414 ATGTCGCTGCAACGCGTTGGAGTAAAGGTCATTCAACCTTTTCTGTTTCATGATAAGCTTT 473  
Db 244 ATGTCGCTGCAACGCGTTGGAGTAAAGGTCATTCAACCTTTTCTGTTTCATGATAAGCTTT 303  
QY 474 TCTGGGATTGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTGTAGCAATGCTAGGCTATG 533  
Db 304 TCTGGGATTGGAGCTGGCAAGAGTTAGCTGAATATGACCTTTTGTAGCAATGCTAGGCTATG 363  
QY 534 TGTATA 539  
Db 364 TATATA 369

Search completed: December 30, 2002, 19:04:44  
Job time : 2235 secs



Db 367 GATGAATGGCTGAATATGACCTCCATCTACAATTGATTTTCATCTTAAGAGAACAGGA 426  
Qy 548 CAGTCCAAAATCTATATCTGGGGATTCACAGGGAACATAAATGGGTTTGGCGCTTTG 607  
Db 427 CAGAAGAAGCTACACTATGTGGCCATTCCCAAGGACACCAATGGTTTGTGCGCTTT 486  
Qy 608 ACGA 611  
Db 487 TCTA 490

## RESULT 2

US-09-186-489-3

; Sequence 3, Application US/09186489

; Patent No. 6375947

; GENERAL INFORMATION:

; APPLICANT: Bolen, Paul L.

; APPLICANT: Cihak, Paul L.

; APPLICANT: Scharpf Jr., Lewis G.

; TITLE OF INVENTION: Purified Recombinant Kid Pregastric Esterase, and

; FILE REFERENCE: 5499/3

; CURRENT APPLICATION NUMBER: US/09/186.489

; CURRENT FILING DATE: 1998-11-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 1411

; TYPE: DNA

; ORGANISM: Kid (Goat)

US-09-186-489-3

Query Match 6.1%; Score 91.2; DB 4; Length 1411;  
Best Local Similarity 54.2%; Pred. No. 2e-14;  
Matches 230; Conservative 0; Mismatches 188; Indels 6; Gaps 2;

Qy 194 GGCTACCCCTGCACCGAGCACAACGTTGAAACAAAGATGGATTCCTTTATCTCTTCAG 253  
Db 191 GGCTACCCCAAGTGAGATGCATAAAGTTATACTGCAGATGCTATATCTTCAGGCTCTAT 250  
Qy 254 CATATCCACATGGCAAAATAAAGCAGCAGAT---AGTACTGGCCCTCCAGTTTCTT 310  
Db 251 CGGATTCCTCATGGAAGAATGATGCTAATCATTTAGCTCAGAGACCTGTGTGTTCTG 310  
Qy 311 CAACATGTCTTTTTCAGGGAGGAGACATGTTTCATAAATCTGCTGAGCAATCACTT 370  
Db 311 CAGCATGTTCTTCTTGCTCAGCTACAACTGGATTTCCACCTTCCCAACAACAGCCTG 370  
Qy 371 GGGTATATCCTTGTGTGATAACGGTTTGTGATTTGGATTGGGAATGTCCTGGGAACGCT 430  
Db 371 GGCTTCCCTGGCAGATGCTGTTTATGACGTGTGGCTGGGGAACAGCAGAGAAACACT 430  
Qy 431 TGGAGTAAAGTCAATTCACACCTTTTCTGTTATGATAGCTTTTCTGGGATTGGAGCTGG 490  
Db 431 TGGGCCCCAGGAACATTTATCTATTTACCTCAACGACTCCCTGAAATTCCTGGGCTTT 490  
Qy 491 CAAGAGTTAGCTGAATATGACCTTTTTCAGCAATGCTAGCTATGCT---ATACAGTCACA 547  
Db 491 GATGAATGGCTGAATATGACCTTCCATCTCAATTTGATTTCAATCTTAAGAGAACAGGA 550  
Qy 548 CAGTCCAAAATCTATATGTGGGGATTCACAGGGAACATAAATGGGTTTGGCGCTTTG 607  
Db 551 CAGAAGAAGCTACACTATGTGGCCATTCCCAAGGACCACTATGGTTTGTGCGCTTT 610  
Qy 608 ACGA 611  
Db 611 TCTA 614

## RESULT 3

US-09-186-489-5

; Sequence 5, Application US/09186489

; Patent No. 6375947

; GENERAL INFORMATION:  
; APPLICANT: Bolen, Paul L.  
; APPLICANT: Cihak, Paul L.  
; APPLICANT: Scharpf Jr., Lewis G.  
; TITLE OF INVENTION: Purified Recombinant Kid Pregastric Esterase, and  
; FILE REFERENCE: 5499/3  
; CURRENT APPLICATION NUMBER: US/09/186.489  
; CURRENT FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 8324  
; TYPE: DNA  
; ORGANISM: Yeast YE-1 expression vector  
US-09-186-489-5

Query Match 6.1%; Score 91.2; DB 4; Length 8324;  
Best Local Similarity 54.2%; Pred. No. 4.4e-14;  
Matches 230; Conservative 0; Mismatches 188; Indels 6; Gaps 2;

Qy 194 GGCTACCCCTGCACCGAGCACAACGTTGAAACAAAGATGGATTCCTTTATCTCTTCAG 253  
Db 1598 GGCTACCCCAAGTGAGATGCATAAAGTTATACTGCAGATGCTATATCTTCAGGCTCTAT 1657  
Qy 254 CATATCCACATGGCAAAATAAAGCAGCAGAT---AGTACTGGCCCTCCAGTTTCTTCTT 310  
Db 1658 CGGATTCCTCATGGAAGAATGATGCTAATCATTTAGCTCAGAGACCTGTGTGTTCTG 1717  
Qy 311 CAACATGTCTTTTTCAGGGAGGAGACACATGTTTTCATAAATCTGCTGAGCAATCACTT 370  
Db 1718 CAGCATGCTCTTCTTGCTCAGCTACAACTGGATTTCCACCTTCCCAACAACAGCCTG 1777  
Qy 371 GGGTATATCCTTGTGTGATAACGCTTTTCTGTTTGGATTGGGAATGTCCTGGGAACGCT 430  
Db 1778 GGCTTCCCTGGCAGATGCTGTTTATGACGTGTGGCTGGGGAACAGCAGAGAAACACT 1837  
Qy 431 TGGAGTAAAGTCAATTCACACCTTTTCTGTTTATGATAGCTTTTCTGGGATTGGAGCTGG 490  
Db 1838 TGGGCCCCAGGAACATTTATCTATTTACAGACTCCCTGAAATTCCTGGGCTTT 1897  
Qy 491 CAAGAGTTAGCTGAATATGACCTTTTTCAGCAATGCTAGCTATGCT---ATACAGTCACA 547  
Db 1898 GATGAATGGCTGAATATGACCTTCCATCTCAATTTGATTTCAATCTTAAGAGAACAGGA 1957  
Qy 548 CAGTCCAAAATCTATATGTGGGGATTCACAGGGAACATAAATGGGTTTGGCGCTTTG 607  
Db 1958 CAGAAGAAGCTACACTATGTGGCCATTCCCAAGGACCACTATGGTTTGTGCGCTTT 2017  
Qy 608 ACGA 611  
Db 2018 TCTA 2021

## RESULT 4

US-09-186-489-4

; Sequence 4, Application US/09186489

; Patent No. 6375947

; GENERAL INFORMATION:

; APPLICANT: Bolen, Paul L.

; APPLICANT: Cihak, Paul L.

; APPLICANT: Scharpf Jr., Lewis G.

; TITLE OF INVENTION: Purified Recombinant Kid Pregastric Esterase, and

; FILE REFERENCE: 5499/3

; CURRENT APPLICATION NUMBER: US/09/186.489

; CURRENT FILING DATE: 1998-11-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1134

; TYPE: DNA

; ORGANISM: Bovine

Matches	228	Conservative	0	Mismatches	190	Indels	6	Gaps	2
Matches	228	Conservative	0	Mismatches	190	Indels	6	Gaps	2

0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Db 127 CGGATTCCATCGGAAAGAAATAATGCTAATCATTTAGGTGAGAGACCTGTTGTGTTCTG 186

QY 311 CAACATGGTCTTTTCAGGAGACACATGGTTCATAAACTCTGCTGAGCAATCACTT 370

DB 18/ CAGCATGGTCTTCTTGGATCAGCCACAACTGGATTTCACACCTGCCCCAAGAACAGCCTG 246

430  
420  
410  
400  
390  
380  
370  
360  
350  
340  
330  
320  
310  
300  
290  
280  
270  
260  
250  
240  
230  
220  
210  
200  
190  
180  
170  
160  
150  
140  
130  
120  
110  
100  
90  
80  
70  
60  
50  
40  
30  
20  
10  
0

Db 307 TGGGGCCAGGAACATTTTATCTATTACCCACACTCCCCCGAATTCTCCCTTTTCACTTTT 366

QY 491 CAAGAGTTAGCTGAATATGACCTTTTAGCAATGCTAGGCTATGTGTATACAGTCAC--A 547

Db 367 GATGAAATGGCGGAATATGACCTTCCATCTACAATTGATTTCACTTAAGGAGAACAGGA 426

QY 548 CAGTCCAAAATTCCTATATGTTGGGGCATTCACAGGGAACATAATGGGTTTGGCGGCTTGG 607

[illegible]

— 2 —

US-09-820-001-1

Patent No. 6387680

; AFFRICANI, MERRILOV, GENNADY et al  
; TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS. NUCLEIC

; TITLE OF INVENTION: THEREOF

; CURRENT APPLICATION NUMBER: US/09/820,001  
 ; CURRENT FILING DATE: 2001-02-28

; SOFTWARE: FastSEQ for Windows Version 4.0

```

; LENGTH: 1360

```

ORGANISM: Human  
IS-09-820-001-1

Query Match 5.5%; Score 81.6; DB 4; Length 1360;

Matches 224; Conservative 0; Mismatches 194; Indels 6; Gaps

194 GGC TAC CCG TGC ACC GAG CAC AAC G T T G A A A C C A A A G A T G G A T T C C T T T T A T C T C T T C A G 253

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

311 CAACATGGCTCTTTTCACGGCAGCAGACACATCGCTTCATTAACATCTCTCCCTCACCAATCACCTT 370

\_\_\_\_\_



```

: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Crissey, Todd M.
: REGISTRATION NUMBER: 37,807
: REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 734 622-7530
: TELEFAX: 734 622-1553
: TELEX:
:

```

```

;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1140 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1140
US-09-073-674-4
Query Match 4.3%; Score 64.4; DB 2: Length 1140;

```

DB	355	TTCTGGGCTTTTCAGCTTTTCACGAGATGGGCTAAATATGACCTTTCCCGCCACCAATTGACTTC	411
QY	531	-ATGTGTATACAGCTCACAGTCCCAAAATTTCTATATGTGGGGCATTCACAGGGAACATATA	589
DB	415	ATCTTGAGAAAACGGACACGACAGCTACACTACGTTGGCCATTCACAGGCGACCAACC	474
QY	590	ATGGGTTTGGCGGCTTTTGACGA	611
DB	475	ATTGGTTTCATCGCCTTTTCCA	496
RESULT 10			
US-08-227-108-6			
; Sequence 6, Application US/08227108			
; Patent No. 5807726			
; GENERAL INFORMATION:			
; APPLICANT: Blanchard, Claire			
; APPLICANT: Benicourt, Claude			
; APPLICANT: Junien, Jean-Louis			
; TITLE OF INVENTION: Recombinant Dog Gastric Lipase			
; NUMBER OF SEQUENCES: 21			

```

RESULT 10
US-08-227-108-6
; Sequence 6, Application US/08227108
; Patent No. 5807726
; GENERAL INFORMATION:
; APPLICANT: Blanchard, Claire
; APPLICANT: Benicourt, Claude
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: Recombinant
; NUMBER OF SEQUENCES: 21

```





[illegible]

```
;
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match      3.1%; Score 46.4; DB 4; Length 1926;
Best Local Similarity 54.0%; Pred. No. 0.012;
Matches 95; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 44 GGCACGCCCCCGGAGGAGCGCTCCCGCTGATCCTCGTCTCTCTGTGCGGTGC 103
Db 377 GTCCTCTCCCGCTCTCGTCTCTCTCCCGCTCTCCCGCTCTCGTCTCTCGTCTC 436
Qy 104 ATCGTCTCCGGAGCGCTCCCGAGCGCGCGCGCTCCCGCGCTCGGTCCGGCTCCGGC 163
Db 437 CTCGCCGCTCTCTCGTCTCTCCCGCTCTCCCGCTCTCGTCTCTCTCCCGCTCTCCCG 496
Qy 164 GGCCTCTGCGACAGCTGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 219
Db 497 GTCCTCCCGCTCTCGTCTCTCTCCCGCTCTCTCCCGCTCTCTCTCTCTCTCTCTCT 552

Search completed: December 30, 2002, 19:06:31
Job time : 102 secs
```

---

```
Qy 413 AATGTCCTGGAACGCGTTGGAGTAAAGGTCATTCACACCTTTTCTGTTTCATGATAAGCTT 472
Db 292 AACACGAGGGGACACACTGGCCGAGGAGTATGTTACTACTCGCCGACTCCGTCGAA 351
Qy 473 TCTCGGGATTGGAGCTGCAAGAGTTAGCTGAATGATGACCTTTTAGCAATGCTAGGCT-- 530
Db 352 TTCTGGGCTTTTCAGCTTTGACGAGATGGCTAAATATGATGACCTTCGCGCCACCATGACTTC 411
Qy 531 -ATGTTGATACAGTCACACACACTCCAAATTTCTATATGTTGGGCGATTTCACAGGGAACATA 589
Db 412 ATCTTTGAAGAAACGGGACAGCAAGTACACTACGTTGGCCATTCACAGGGCACCACC 471
Qy 590 ATGGGTTTGGCGGCTTTTGACGA 611
Db 472 ATGGGTTTCATCGCCTTTTCCA 493

RESULT 14
US-09-820-001-3
; Sequence 3, Application US/09820001
; Patent No. 6387680
; GENERAL INFORMATION:
; APPLICANT: MERRULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN LIPASE PROTEINS, AND USES
; FILE REFERENCE: CL001186
; CURRENT APPLICATION NUMBER: US/09/820,001
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 22067
; TYPE: DNA
; ORGANISM: Human
US-09-820-001-3

Query Match      3.5%; Score 52.2; DB 4; Length 22067;
Best Local Similarity 53.7%; Pred. No. 0.001;
Matches 108; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 288 GTACTGGCCCTCCAGTTTCTTCAACATGGTCTTTTTCAGGAGGAGACACACATGGTTCA 347
Db 6387 GTCCCGCCGCAAGTTGTTGATATGCAGCATGCCCTGTTTGCAGACAAATGCCCTACTGGCTTG 6446
Qy 348 TAACTCTGCTGAGCAATCACTTGGGTATATCCTTGCTGATGATAACGGTTTGTGTTTGGGA 407
Db 6447 AGAATTATGCCAATGGAAGCCTTGGATTCTCTAGCAGATGCAGGTTTATGATGATGGA 6506
Qy 408 TTGGAATGTCCTGGAACGCGTTGGAGTAAAGGTCATTCACACCTTTTCTGTTTCATGATA 467
Db 6507 TGGGAACAGTCGGGGAACACTTGGTCAAGACACACAAACACTCTCAGACAGAGATG 6566
Qy 468 AGCTTTTCTGGGATTGGAGCT 488
Db 6567 AGAAATTCGGGCCCTTAGGT 6587

RESULT 15
US-09-249-585A-4
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
```